The Reference Problem and How Children Use Gesture and Grammatical Number to Solve It

Emma Healey

PhD - The University of Edinburgh - 2019

I confirm that this thesis, presented for the degree of PhD, has

i) been composed entirely by me
ii) been solely the result of my own work
iii) not been submitted for any other degree or professional qualification

Emma Healey

Contents

Table of Figures

Table of Tables

Acknowledgements

The work presented in this thesis was fully funded by the Carnegie Trust. I am grateful for their financial support and for their understanding and flexibility, which was equally as important to my success. This project would not have been possible without them.

To all the parents and children who took part in these experiments, thank you so much for volunteering. This type of research would not be possible without your willingness to be involved. As researchers we are indebted to you and hope you and your children enjoy taking part. Thanks also Families magazine for allowing me to advertise my experiments free of charge. I would also like to thank Lara Arnason, PhD, for acting in the videos. Her child-friendly voice and demeanour were instrumental in the success of my experiments and only one child, out of more than a hundred, was visibly afraid.

Thanks also to my supervisors Barbora Skarabela and Mits Ota. I owe much of my academic training to them: Barbora has been my supervisor since I wrote my honours dissertation and Mits since I wrote my MSc thesis. I have been inspired by their knowledge of the field, their intellectual creativity and their academic rigor. I am grateful to them for their consistent understanding, support and engagement. They are a great team. All mistakes are my own.

I have been a member of the department of Linguistics and English Language at the University of Edinburgh since 2003 and it has been an incredibly nurturing and supportive research environment. I would particularly like to thank Katie and Toni from the postgraduate office for their enthusiasm, encouragement and provision of chocolate. I would also like to thank Alisdair Tullo for his help with R and for being an encouraging and patient teacher. Formerly of the department, I would also like to thank Lynsey Buchanan, Ziggy Campbell and Eddie Dubourg.

Thanks are also due to my family, especially my husband, Luke; my mother, Karen; my father, Ross and my brother, Dan. They have been such a wonderful source of love and support throughout this project and beyond and I am incredibly thankful for them. Likewise, I would like to thank my friends, too many to name, who continue to inspire and encourage.

Finally, I would like to say how thankful and grateful I am to God, without whom my life would be very different and none of this would have been possible.

Lay Summary

This thesis considers the reference problem: On hearing an unknown word how does a listener know what that word refers to? This is the task faced by children as they acquire their first language(s). Previous research has shown that children use many different strategies to solve the reference problem. In two key strategies, children employ gesture and grammatical number. When using gesture, children associate novel labels with objects speakers look at or point to. When using grammatical number, children associate plural labels with groups of objects and singular nouns with individual objects. The present research looks in more detail at these processes and asks three questions. Experiment 1 asks whether 2- and 3-year-old children are more attentive to gestures that are made up of body orientation and pointing than to gestures that involve body orientation only. Experiment 2 asks whether children of this age are more attentive to sentences that contain number marking on the noun and verb than to sentences that contain number marking on the noun only. Experiment 3 takes these cues together and asks whether children are more attentive to body orientation and number marking combined than to either body orientation or number marking on its own. The results suggest that children's behaviour in these tasks depends on the specific methodology used and that previous studies may overestimate children's abilities in these areas.

Abstract

This thesis examines 2- and 3-year-olds' understanding of social cues (specifically body orientation and pointing) and grammatical cues (specifically number marking) to reference in a series of three experiments using the Intermodal Preferential Looking Paradigm.

Experiment 1 examines grammatical cues. It investigates whether 2- and 3-year-olds can follow grammatical number when the potential referents belong to the same category (e.g. one knife guard versus 14 knife guards). All previous relevant studies have involved potential referents that belong to different categories (e.g. one lemon juicer versus 14 honey spoons). It is possible that this manipulation makes the task of following grammatical number easier as the categorisation component is already done. If so, previous studies may overestimate children's ability to follow grammatical number in the real world. In addition, Experiment 1 asks whether children follow the combination of nominal and verbal number marking (e.g. 'There are the *zoots*!') more than they follow nominal number marking alone (e.g. 'Look at the *zoot*!'). One previous study shows a preference for multiple number marking, but only with familiar labels and referents. A preference of this kind is referred to as the effect of cue quantity.

Experiment 2 examines social cues and asks whether 2- and 3-year-olds can follow body and head orientation in a referential context. This has not previously been tested experimentally and has implications for children's ability to learn from observation. Experiment 2 also asks whether children follow the combination of body orientation and pointing more than body orientation alone. In the existing literature there are hints of this preference, but previous studies are inconclusive. The experiment therefore looks for the effect of cue quantity in both grammatical and social cues.

Experiments 1 and 2 were designed as pre-tests for Experiment 3 to demonstrate that children could follow the targeted cues (body orientation and number marking) in isolation. Experiment 3 builds on these experiments by exploring changes in children's use of body orientation and number marking across development. In doing so, it aims to bring evidence to bear on a possible shift from reliance on social cues to a reliance on grammatical cues. This is achieved by comparing 2- and 3-year-olds' use of these cues in two conditions: a congruent condition (in which they point to the same referent), and an incongruous condition (in which they point to different referents).

Experiment 1 showed that 2- and 3-year-olds did not follow grammatical number when the potential referents belonged to the same category. Experiment 2 showed 2- and 3-year-olds did not follow head and body orientation. Experiment 3 showed, however, that when these cues were presented together (i.e. with both pointing at the same target) children did show some sensitivity to these cues. This was only the case on singular trials (i.e. when the speaker turned to look at the individual object and produced a sentence containing a singular noun). I argue that in this condition perceptual saliency, as well as body orientation and grammatical number, pointed to the individual object being the target. I interpret this either as evidence in support of the effect of cue quantity or as evidence that 2-year-olds, but not 3-year-olds, continue to use perceptual salience as a cue to reference.

Chapter 1 Introduction

1.1 The Reference Problem

Quine (1969) famously posed the reference problem: On hearing an utterance in an unknown language how does the listener know what that word refers to? The answer is that they can never be sure. In the classic statement of the problem, a linguist travels to a foreign land and sees a local pointing to a rabbit and exclaiming '*Gavagai*!' From this information alone, the possible meanings of this utterance are limitless: 'Look!', 'Danger!', 'Dinner!', 'Rabbit!' are all plausible candidates and the list could go on indefinitely. Having spent some time in the company of native speakers, the traveller will acquire some understanding of the language and culture and will be able to make a more informed guess about the likely meaning of statements in context. However, even if she is able to identify a particularly good candidate (such as 'rabbit'), she is still faced with the problem that the meanings of words change across contexts: What is referred to as 'rabbit' today may be a different thing from what is referred to as 'rabbit' tomorrow.

In this way, the reference problem is not confined to theoretical travellers in unfamiliar lands, or even to children acquiring their native language. Rather, it is a problem that language users are faced with whenever they use language. Imagine, for example, as an adult native-speaker of a given language you hear a sentence like 'The goat's escaping!' Using your knowledge of your native lexicon and morphological system, you can infer that a bearded, clambering animal is currently in the process of evading their captivity. You do not, however, know which one, of all the goats, is 'the goat'. While a speaker can understand the meaning of a sentence, they still have this referential problem to solve. The reference problem can therefore be summarised as follows: How do listeners know what speakers are talking about?

The simple answer is that listeners use many different types of information to work out the likely meanings of words and phrases in context. For example, one way of solving the reference problem is to pay attention to speakers' verbal and non-verbal behaviour and use this information to infer their communicative intentions. As noted above, a speaker may simultaneously point to an object or event while linguistically describing it ('Look! The cow and the pig are dancing!'). By following this gesture, the listener can infer that the speaker is likely to be referring to some aspect of the targeted scene. This does not solve the problem entirely, of course (the speaker could be talking about the pig, the cow, the dance…), but it does allow the listener to limit the hypotheses they entertain (the speaker is unlikely to be talking about the chicken and the duck who are arguing on the other side of the room, for example). Over the course of many such episodes, the listener might notice a correlation between hearing ‘pig’ and seeing a curly-tailed, snout-nosed beastie and come to form a referential link between the two. Having done so, they may realise that when there is only one such beast they hear ‘pig’, but when there is more than one they hear ‘pigs’. Likewise, when they hear ‘cow’ there is just one shaggy, horned creature present, but when they hear ‘cows’ there are many. Consequently, they could infer that if a speaker wants to refer to a group of things they add ‘s’ to the label they use to refer to just one of that thing. Over the course of time, a listener can therefore use a combination of statistical cues, social cues and linguistic cues to make inferences about communicative intentions and about the meanings of words and phrases.

There is consensus among researchers in this area that children use many different types of information to learn the meanings of words (for summaries see Bloom, 2000; Rowland, 2014). The differences between theoretical accounts lie in which types of information are considered primary, and which secondary. For example, some researchers highlight the role of domain-neutral associative mechanisms and argue that statistical learning is at the heart of lexical acquisition (e.g. Smith and colleagues, see for example Samuelson and Smith (1998)). Others emphasise the role of the learner’s emerging socio-cognitive awareness and argue that intention-reading and role reversal imitation guide the learning process (e.g. Tomasello and colleagues, see for example Akhtar and Tomasello (2001)). A third group draws attention to linguistic context as a source of information and suggests the process is largely facilitated by the learner’s use of grammatical cues (e.g. Gleitman and colleagues, see for example Gleitman, Cassidy, Nappa, Papafragou, and Trueswell (2005)). Once a learner has acquired some words using these methods they may go on to develop specific word learning heuristics that will help them to learn more words (e.g. Hirsh-Pasek, Golinkoff, & Hollich, 2001)). One way of integrating these cues is to consider the issue developmentally. The cues that learners have available to them are different at different points in development. For example, linguistic cues (including lexical and grammatical cues) are only available after the learner has had some exposure to the language they are acquiring. As such, the cues used in early development are necessarily non-linguistic. Indeed, in the earliest stages of lexical acquisition (around 10 months old), learners tend to rely on perceptual cues such as salience and associate newly heard novel labels with perceptually salient objects (Pruden, Hirsh-Pasek, Golinkoff, & Hennon, 2006). Across the course of many naming episodes, infants notice that speakers will often label things that they are looking at, pointing to or handling and learn that these kinds of behaviours are reliable indicators of communicative intent. One way to work out what a word means is therefore to consider what the speaker is paying attention to when that word is used. Although they can follow these kinds of cues far earlier in development in simple attention-only tasks (see Section 1.2 below), learners are able to make these kind of referential inferences from the age of 15 months (Houston-Price, Plunkett, & Duffy, 2006).

Social cues are not limited to ostensive behaviours of this kind. By the age of 2 years, children are able to use more subtle indicators of a speaker's communicative intention to guide their referential choices. These types of indicators include facial expression, prosody and tone of voice (Tomasello, Strosberg, & Akhtar, 1996) as well as the wider social context. For example, 2-year-olds understand that previously unseen objects are more likely to elicit an excited response from a speaker than previously seen objects (Akhtar, Carpenter, & Tomasello, 1996). By the time they are two years old, children have learned that social cues, in particular gaze direction, are more reliable than perceptual cues: They will follow the former over the latter when the two conflict (Hollich et al., 2000). This behaviour represents a developmental shift from a reliance on perceptual cues to a reliance on social cues. The question remains open whether perceptual cues, while less useful than social (and other) cues, remain a source of information for learners as they develop, or whether they are supplanted entirely by social (and other) cues.

The process by which learners preferentially use the cues which have been most useful in the past is known as Guided Distributional Learning (Golinkoff & Hirsh-Pasek, 2006; Hirsh-Pasek & Golinkoff, 2008; Hollich et al., 2000). As Paquette-Smith and Johnson (2016) have noted, this mechanism predicts a second developmental shift, this time from reliance on social cues to reliance on grammatical cues. I elaborate on this below.

Social cues are not always available as a speaker will not always be looking at, pointing to or handling an object she is talking about. As conversations move away from the here and now, grammatical cues become more relevant. Grammatical cues are wholly consistent. To take grammatical number as an example, a group of objects of like-kind will necessarily be referred to using plural morpho-syntax. Therefore, if a learner hears a novel label in the context of plural morpho-syntax, they know they are looking for a group. This understanding is reinforced every time a speaker refers to a group that is physically present. Over time, children may learn that this is an extremely reliable cue to reference and they may therefore come to weight grammatical cues more heavily than social cues. There is evidence that this is the case in the acquisition of verbs (Nappa, Wessel, McEldoon, Gleitman, & Trueswell, 2009). One of the main aims of the current thesis is to explore whether a shift of this kind is evident in children's understanding of novel nouns. Grammatical cues specify relationships between arguments and are therefore necessary for the acquisition of verbs in a way that they are not necessary for the acquisition of nouns. This alone is enough to suggest that grammatical cues might be weighted less heavily in nominal acquisition than in verbal acquisition. Furthermore, while the claim that cues like grammatical number marking are wholly consistent might be true in theory, it is not necessarily the case in practice. As Lukyanenko and Fisher (2016) point out, number mismatches such a 'Where's the apples?', where a singular verb co-occurs with a plural noun' are increasingly common in American English. Several conversations overheard during the writing of this thesis, including 'Is the kids' cups and plates still available?' suggest, at least anecdotally, that this is also the case for British English. As such, the mapping between form and meaning may, in some cases at least, be less consistent than first appears. If children are sensitive to this kind of statistical (ir)regularity they may therefore not learn to weight grammatical cues more heavily than social cues.

To increase knowledge in this area, this research examines how children use and understand social cues and grammatical cues. In addition to tracking developmental changes in the weightings of these cues, it considers the role of what I have called 'cue quantity' and assesses the possibility that young children will be more responsive to social and grammatical cues that contain more pieces of relevant referential information. This is discussed in more detail below. In the remainder of this chapter, I discuss children's ability to use social cues and grammatical cues and explore the possibility that 2- and 3-year-olds may follow gestures and sentences with higher cue quantity more than those with lower cue quantity. I also introduce two experiments exploring these issues. I begin by considering the role of social cues.

1.2 Solving the Reference Problem Using Social Cues

This section discusses children's ability to solve the reference problem using social cues. The first subsection gives an overview of social cues and explains how children can use this type of information to learn the meanings of new words. The second subsection then examines two of these cues, gaze direction and pointing, in more detail and explains how children show sensitivity to these cues in non-linguistic tasks. The third subsection looks again at these two cues, considering from what point children can use them to learn words and how this ability changes across development. The final subsection identifies some gaps in our understanding and introduces an experiment designed to address them.

1.2.1 Social cues and word learning

As described in Section 1.1, children are able to use behavioural cues from speakers to make inferences about their referential intentions. These behaviours are known as social cues and there is a wide body of literature demonstrating children's ability to use these cues to inform their interpretations of novel labels (for overviews see Bloom, 2000; Rowland, 2014). Children's ability to learn the meanings of new words is therefore facilitated by their emerging socio-cognitive awareness and their ability to understand other people as intentional agents (Akhtar & Tomasello, 2001; Tomasello, 2003). To illustrate, children first use their intention-reading skills to discern how a speaker is using a particular label to direct the attention of the listener. They then use role-reversal imitation to work out that they themselves can use the same label to direct the attention of others to that that same referent. On this account, a child who acquires a new label acquires not an arbitrary mapping between label and referent, but rather 'a communicative device understood intersubjectively from both sides of the interaction' (Tomasello, 1999, p. 106; 2001). For the purpose of this thesis, I am interested primarily in the first stage of this process: children's ability to discern the communicative intentions of their interlocutors. Conversations with young children tend to be grounded in the here and now. Deictic cues such as social eye-gaze and pointing are therefore extremely helpful in facilitating children's understanding of referential intent (see below). In the following subsection we consider the developmental trajectories of children's sensitivity to these cues. I begin by considering how children use gaze direction and pointing to follow attention in contexts that do not involve language, before going on to discuss how and when they use these cues referentially.

1.2.2 Gaze following and pointing in non-linguistic tasks

Infants follow gaze direction robustly and reliably by the age of 9 months (Houston-Price et al., 2006; Rowland, 2014; Stephens & Matthews, 2014), although there is some evidence that younger children do so also (Butterworth & Grover, 1988, 1990; D'Entremont, Hains, & Muir, 1997; Morales, Mundy, & Rojas, 1998). At this age, though, they tend to follow the direction of the head, rather than the eyes (Moore, 2008). Evidence for this comes from the fact that they are equally likely to follow the gaze direction of a speaker who turns her head with her eyes closed or blind-folded, as a speaker who turns her head with her eyes open and not blind-folded (Brooks & Meltzoff, 2002, 2005). This suggests they have not yet learned that in order for them to see, speakers need to have their eyes open and unobscured. By the time they are 10 to 11 months old, however, children understand the importance of the eyes and are more likely to follow the gaze direction of speakers whose eyes are open than the gaze direction of speakers whose eyes are closed (Brooks & Meltzoff, 2005).

Infants' tendency to identify and follow speakers' gaze direction increases between the ages of 9 and 15 months (Carpenter, Nagell, & Tomasello, 1998). Indeed, 12-month-old infants regard gaze direction as such a useful source of information about other people and their attentional states, that if they recognise a speaker can see something that they themselves can't, they will move in order to be able to see it (Moll & Tomasello, 2004). By the time they are 12 months old, children understand the correlation between gaze direction and head/body orientation. Although they understand that in order to see, a speaker needs to have their eyes open and unobscured (see above), they also understand that if a speaker's eyes are not visible they can use her head and/or body orientation to infer her gaze direction. Tomasello, Hare, Lehmann, and Call (2007) provide evidence of this. In their experiment, the experimenter sat with his back towards the participants and looked up towards the ceiling. Both 12- and 18-month-olds looked at the ceiling more often in this condition than when the actor faced the child and did not look at the ceiling, with the 18-month-olds doing so more often than the 12-month-olds.

Deák, Flom, and Pick (2000) provide supporting evidence that 18-month-olds can infer gaze direction on the basis of head and body orientation. In this experiment, parents looked at a target that was either in front of the child or behind the child. The results showed that 18-month-olds followed their parent's gaze and attended to the target both when the target was in front of them and when it was behind them. Importantly, when the parent turned to look at a target in front of the infant, the infant could not see their parent's eyes and had to infer their gaze direction from the orientation of their head. These results therefore indicate that from the age of 12 to 18 months, children understand that the orientation of a speaker's head provides information about their gaze direction. This point is returned to later.

Returning to the developmental trajectory of gaze following, by the time they are 14 months old, infants understand that being able to see is not enough for a speaker's gaze direction to be informative. Rather, for gaze direction to be informative, speakers have to be able to see something in particular. Specifically, 14-month-olds follow a speaker's direction of gaze to a target when there is no barrier in between the gazer and the target and also when there is a clear barrier, but not when there is an opaque barrier (Dunphy-Lelii & Wellman, 2004). This suggests that by this age, infants understand that gaze direction is referential, and only follow a speaker's direction of gaze when she is able to see something. Infants' ability to follow gaze direction is not, therefore, automatic, but is rather based on their emerging understanding of other people as intentional agents.

Fifteen- to 18-month-old infants respond more to the combination of head and eye orientation than to eye orientation alone (Corkum & Moore, 1995). In particular 15- to 18-month-olds successfully follow the attention of another when it is signalled by a change in the orientation of the eyes and the head, but not when it is signalled by a change in orientation of one or the other (Corkum & Moore, 1995). Butterworth and Jarrett (1991) provide converging evidence with 18-month-olds and Lempers (1979) with children as young as 11 months old. Both these studies, however, had some methodological flaws which make interpretation of their results difficult. What might explain this behaviour? One explanation is that between the ages of 15 and 18 months, infants interpret a shift of head and eye orientation as a more emphatic gesture than a shift of either in isolation. An alternative explanation is that a shift of head and eye orientation is more salient than a shift of either in isolation. I return to this point later.

With regard to pointing, there is some evidence that production precedes comprehension. Specifically, in terms of comprehension, i.e. interpreting other people's pointing gestures, 6- to 9-month-olds focus on the finger and hand when a person points and it is not until the age of 12 months that infants reliably follow a point to an intended target (Stephens & Matthews, 2014). In terms of production, though, infants begin to make pointing gestures between the ages of 9 and 12 months (Golinkoff & Hirsh-Pasek, 2000). During this time, pointing gestures can be used to request objects, to share interest with others or to direct attention to people, objects and locations (Ng, Demir, & So, 2015). As was the case with gaze following, there is evidence that infants' attention to pointing increases between the ages of 9 and 15 months (Carpenter, Nagell, et al., 1998).

Linking in with our earlier discussion of gaze direction, there is evidence that 18-month-olds are more attentive to pointing when it is accompanied by another gesture than when it is presented in isolation. In particular, Deák et al. (2000) show that 18-month-olds respond more to gaze direction and pointing than to gaze direction alone. As was the case with children's preference for head movement coupled with eye-movement, there are two explanations for this: emphatic-ness and saliency. Furthermore, Deák et al. (2000) showed that infants looked longer at the target following a larger head turn compared with a smaller head turn. One explanation for this is that the larger head turn is a more deliberate action and a clearer indicator of intention than the smaller head turn. Another explanation is that a larger head turn is more perceptually salient than a smaller head turn. Exploring children's preference for 'heavy' gestures is one of the aims of the present thesis and so it is considered in more detail below.

The previous sections discussed how presenting multiple social cues (e.g. gaze direction versus gaze direction and pointing) might have a facilitative effect on young learners' tendency to follow those cues. In the following sections we will discuss how this is also the case for multiple grammatical cues (e.g. nominal number marking versus nominal and verbal number marking, see Section 1.3) and also when cues of different types combine (e.g. when perceptual cues and social cues or social cues and grammatical cues are presented in alignment, see Section 1.4). From now on, this will be referred to as the effect of cue quantity. To elaborate, a cue that contains more than one cue of the same type (e.g. perceptual, social or grammatical) or that is formed from two cues of different types (perceptual and social, social and grammatical etc.) has a higher cue quantity than a cue that contains only one cue. In this way, cue quantity pertains to the number of cues.

As noted above, there are two reasons why children might follow multiple cues over single cues (a) Multiple cues are more perceptually salient than single cues; (b) Multiple cues are more emphatic than single cues. To elaborate, in the case of social cues the combination of looking and pointing could draw more attention because it is more visually salient or because it is indicative of greater intentionality on the part of the gesturer. In the case of grammatical cues, the combination of nominal and verbal number marking could draw more attention because it is more audibly salient. The question of whether it could be considered more emphatic is trickier: If a speaker wanted to emphasise the plurality or singularity of a referent would she use more number marked elements? Or is the decision to say 'Look at the rabbits!' or 'There are some rabbits!' guided by other factors? This is not a question that can be answered intuitively. One way of bringing some evidence to bear on this issue is to consider the issue developmentally.

In particular, as discussed earlier, there is evidence that the relative perceptual salience of the potential referents affects younger learners' referential decisions to a greater extent than it affects those of older learners (Hollich et al., 2000; Pruden et al., 2006). It is possible that this trend will extend to the perceptual salience of the cue itself, and that younger learners will be more affected by the salience of the cue than older learners. In this way, if the effect of cue quantity is attributable to perceptual salience, one would expect to see the effect most strongly in younger learners. Alternatively, if the effect of cue quantity is attributable to increased intentionality, one would expect to see the effect more strongly in older learners. This is because older learners are better at intention-reading than younger learners. For this reason the experiments presented here will test children across a wide age-range, namely 2- and 3-year-olds.

In this section thus far, I have shown how infants use gaze direction and pointing to follow the attention of others and have provided some evidence in support of the effect of cue quantity. In the following section I build on this and discuss how infants and young children use these gestures to make referential choices and identify the likely referents of novel labels. As I will show, cue quantity also has a role to play here. I begin by considering the role of gaze direction.

1.2.3 Gaze following and pointing in referential tasks

In the earliest stages of lexical acquisition (around 10 months of age) infants do not use gaze direction to inform their referential choices and instead make referential choices based on the relative perceptual salience of the potential referents (Pruden et al., 2006). By the time they are 15 months old, however, infants do use gaze direction to inform their interpretation of novel labels and associate novel labels with the focus of a speaker's gaze (Houston-Price et al., 2006, Experiment 1). By the age of 18 and 19 months (but not 16 and 17 months) they treat the speaker's gaze direction as such a reliable cue to reference that they associate novel labels with the focus of the speaker's gaze even if they themselves were looking at something else when the label was introduced (Baldwin, 1991, 1993) Additionally, Hollich et al. (2000) show that by the age of 24 months, children associated novel labels with gazed-upon objects even if the gazed-upon object is less perceptually salient than a competing potential referent. There is therefore evidence that by the age of 24 months children understand gaze direction to be an extremely useful source of information for word learning.

It is important to note that this is not just an associative mechanism. As Grassmann (2014) asserts, children do not 'blindly' associate novel labels with gazed-upon objects, but rather they do so only when the wider social context suggests that gaze direction is a reliable indication of communicative intention. This is the case from as young as 14 months of age. Specifically, Chow, Poulin-Dubois, and Lewis (2008) showed that 14-month-olds do not follow a speaker's gaze direction if that speaker has previously shown their direction of gaze to be an unreliable indicator of their intentional state (see also the barrier studies described above). Similarly, Grassmann, Stracke, and Tomasello (2009) show that older children (between the ages of 2;9 – 2;11) do not use gaze direction as a cue to reference when the speaker excitedly produces a novel label, but is looking at something that she has seen before, i.e. something that is 'old news' (Grassmann, 2014). Similarly, Nurmsoo and Bloom (2008) found that 3- and 4-year-olds use gaze direction as a cue to reference after hearing sentences such as ‘There’s the *nurmy*!’, but not when they heard sentences like ‘Where’s the *nurmy*?’. This suggests that they understand that a speaker is likely to be looking at the intended referent in the former case, but not the latter. Studies such as these suggest that children use social eye-gaze as a cue to reference only when it is informative about a speaker’s referential intentions.

With regard to pointing, children associate novel labels with pointed-at objects by the age of 18 months (Briganti & Cohen, 2011). This is not the case at 14 months. The ability to use pointing as a cue to reference therefore emerges at around the same time as the ability to use gaze direction as a cue to reference. To our knowledge, Briganti and Cohen (2011) is the only study to explicitly test children’s ability to use pointing gestures to identify the referents of novel labels. There are some limitations with this experiment, however. The first limitation is that participants’ looks-to-target were compared to chance, but it is not clear from the write-up how chance was calculated. The second limitation concerns the articulation of the actor’s gesture. In particular, the write-up states that the videos showed that the actor was ‘pointing and nodding…towards two separate novel moving objects’. This suggests that the speaker may also have been looking at the target while she pointed. As such, the participants’ behaviour could be attributed solely to their attention to gaze direction or to the combination of gaze direction and pointing. This is an interesting possibility as, in non-linguistic, attention-only tasks, infants respond more to gestures comprising multiple elements, as we have already seen (Briganti & Cohen, 2011). Below I discuss evidence that this is also the case for linguistic tasks.

There is evidence that children’s referential decisions are informed by looking and pointing gestures in contexts where they are not informed by pointing alone. To illustrate, it is well established that in referent selection tasks, learners tend to select objects they do not already have names for as the intended referents of novel labels (Merriman & Bowman, 1989). This is the case from approximately 17 months of age (Halberda, 2006). Jaswal and Hansen (2006) provide evidence to suggest that for 3- and 4-year-olds this behaviour is robust even if the speaker gestures towards a familiar object as she makes her request. In this study, participants were shown a familiar object (such as an elephant) and an unfamiliar object (such as a honey spoon) and asked ‘Can you give me the *blicket*?’ At the same time, the experimenter either looked at or pointed to the familiar object. In looking-only trials, participants selected the unfamiliar object 88% of the time. In the pointing-only trials, this was the case 75% of the time. In both cases, the participants therefore ignored the gesture (and the information it provided about the speaker's communicative intention) and instead made their referential decisions based on their own assumptions about the likely meanings the newly-heard label.

A follow-up study, however, showed that if the experimenter looked at *and* pointed to the familiar object the pattern was reversed (Grassmann & Tomasello, 2010). This study used the same basic procedure as described above: Two- and four-year-old participants were shown a familiar object and an unfamiliar object and were asked to 'Find the *baffe*!' In one condition the experimenter pointed at the familiar object with her index finger while looking at the child (pointing only condition). This was comparable to the pointing gesture used in the original study. In the second condition, the experimenter pointed at the familiar object with her forearm and index finger and alternated her gaze between the familiar object and the participant (looking and pointing condition). In the pointing only condition, participants selected the novel object on 55% of trials. In the looking and pointing condition, however, they selected the unfamiliar object on only 2% of trials. In this case, learners used the gesture to inform their referential choices and override their tendency to associate novel labels with novel objects. This suggests that in some contexts children will follow gestures comprising two elements (e.g. looking and pointing) when they do not follow gestures comprising a single element (e.g. pointing only).

For the present purposes, one issue with Grassmann and Tomasello (2010)'s study is that it conflates the *amount* of referential information a cue provides with the *quality* of the information it provides. In particular, the pointing only gestures were produced in less facilitative social contexts than the looking and pointing gesture. That is, the experimenter did not smile and her demeanour was less friendly. This alone could explain participants' lack of engagement. In addition, the pointing gesture itself was articulated differently in the pointing only scenarios (in which it consisted of an extended index finger) and in the pointing and looking scenarios (in which it consisted of an extended index finger and raised forearm). In this way, the comparison was between a small pointing gesture in an unsupportive social context and a big pointing gesture coupled with looking in a supportive social context. In this thesis I aim to tease apart these differences and ascertain whether learners' referential choices are more guided by looking and pointing gestures than looking or pointing gestures when the social context and the articulation of the gestures are controlled for. This is discussed in more detail in the following section.

1.2.4 Introducing an experiment exploring children's understanding of social cues

Chapter 4 presents an experiment designed to test whether 2- and 3-year-old children follow social cues comprising two components (e.g. looking and pointing) more than they follow either one of those components in isolation. I reason that, in the real world, children are more likely to experience natural looking gestures in the absence of pointing than they are to experience pointing gestures in the absence of looking. I therefore compare looking and pointing with looking only. As noted above, there is some evidence that in simple attention tasks children are more responsive to gestures comprising two elements than those comprising one. The only studies with a referential component that touch on this question, however, have an experimental confound and so this experiment will bring new evidence to bear on children's cue preferences in referential tasks.

My experiment tests children from across a broad age range for two reasons. The first is because developmental differences may be enlightening with regard the origin of any observed preference for the two element gesture. To elaborate, as discussed in Section 1.2.2 above, a preference for a gesture comprising two elements could be explained either by the fact that a two-element gesture is indicative of greater pragmatic intentionality or because it is more perceptually salient. If the former is the case, and children pay more attention to two-element gestures because they have learned that two-element gestures are better indicators of the gesturer's communicative intention, then one should expect older children to display this preference to a greater extent than younger children. This is because older children are better intention-readers than younger children.[1] It is not clear what, if any, developmental differences we should expect if a preference for the two-element gesture is due to its perceptual salience. In particular, in word learning studies (e.g. Hollich et al., 2000; Pruden et al., 2006) the role of perceptual salience decreases across development and it is possible that we will see a similar pattern here. Alternatively, no developmental differences may be found.

A second reason for looking at 2- and 3-year-olds is that this thesis will also present an experiment in which we compare children's use of social and grammatical cues across development (Experiment 3). In particular, it will explore the possibility of a shift in reliance on social cues to a reliance on grammatical cues across this period (i.e. from the earliest point at which they can use both cues). Before looking for a preference between cues, it is necessary to demonstrate that children of the targeted age are actually able to use both cues. In this way, our experiment with social cues (Experiment

[1] Unless children across the range tested reach ceiling levels

2) acts as a pre-test for Experiment 3. Experiment 1 (discussed in the following section) performs the same function with regard to children's understanding of the grammatical cue.

Another gap in our understanding of children's understanding and use of gaze direction concerns their ability to infer gaze direction from the orientation of the head and the body. As described above, there is some evidence from non-linguistic, attention-only tasks that children can make this inference from the age of 12 months. There is, however, no existing evidence that this is the case in tasks with a referential component. This is an important question because, according to one influential theory of language learning, an important part of language acquisition is a learner's ability to learn from 'overseen' interactions, i.e. interactions that they are not directly a part of (Tomasello, 2003). In interactions of this kind, it is likely that a learner will not always be able to see the speaker's eyes clearly and as such gaze direction may have to be inferred on the basis of body orientation. If children can learn from overseen interactions, we should expect them to be able to infer a speaker's gaze direction from her body orientation.

The experiment presented in Chapter 4 therefore seeks to answer the following questions:

1. Are 2- and 3-year-olds more responsive to gestures comprising multiple elements? Does this change across development?
2. Can 2- and 3-year-olds infer gaze direction on the basis of head and body orientation?

The preceding section has looked at the role of social cues in children's acquisition of novel labels and has considered their understanding and use of two such cues across development. The following section considers the role of grammatical cues.

1.3 Solving the Reference Problem Using Grammatical Cues

The following sections discuss children's ability to solve the reference problem using grammatical cues. The first section gives an overview of the grammatical cues that exist in the environment and explains how children can use them to learn the meanings of words. The second section presents evidence in support of children's ability to use these cues. The third section considers one of these cues, grammatical number marking, in more detail. The final section identifies some gaps in our understanding and introduces an experiment designed to address them.

1.3.1 Grammatical cues and word learning

In natural language there are correlations between form and meaning. That is, certain meanings are expressed using particular forms. For example, in English transitive verbs tend to refer to causative actions and intransitive verbs to non-causative actions. A child learning English can therefore infer that if a novel label is introduced as a transitive verb (e.g. ‘The duck is *glorping* the bunny!’) it refers to a causative event and if it is introduced as an intransitive verb (e.g. ‘The duck is *glorping*!’) it refers to a non-causative action. Other regular form-to-meaning mappings that learners could exploit in the interpretation of novel verbs include ditransitive verbs and the transfer of possession (e.g. ‘The duck is *glorping* the bunny the carrot!’) and verbs with clausal complements and mental states (e.g. The duck *glorps* that the bunny is hungry). This is not applicable only to verbs. Another regular pattern that English learners could use is between adjectival morpho-syntax (e.g. ‘a *zav* one’, ‘a *zavy* one’ or a ‘a *zavish* one’) and properties.

Similarly, with regard to interpreting nouns, English morpho-syntax distinguishes between count nouns which refer to countable, individuated entities, i.e. things, and mass nouns which refer to non-countable, non-individuated entities, i.e. substances. A child learning English can therefore infer that if a novel label is introduced with count noun morpho-syntax (e.g. ‘This is a *zav*!’) it refers to a countable object, whereas if it is introduced with mass noun morpho-syntax (e.g. ‘This is some *zav*!’) it refers to a substance. In addition, in English proper nouns are marked by zero inflection and occur without an article (e.g. ‘This is *Zav*’). Learners who hear a novel label introduced in this way can therefore infer than it refers not to a category (like common nouns do), but rather to an individual. Another class of nouns is collective nouns like ‘flock’ or ‘herd’ which refer not to individuals, but to groups of individuals of the same kind. Syntactically, they are referred to in the same way as singular count nouns (compare ‘There is a flock’ and ‘There is a sheep’). If, however, the singular count noun syntax is used in relation to a group (as it would be in the first example), learners might use the mismatch between form and meaning to infer that the label refers to a group. Finally, like many languages, English linguistically encodes the semantic difference between one and more than one. This means that individuals are referred to using singular morpho-syntax (e.g. ‘There is a *blicket*’) and groups are referred to using plural morpho-syntax (e.g. ‘There are some *blickets*’). As shown in these examples, in English verbs, nouns and determiners are all number marked. In languages such as Spanish and Italian this is also the case for adjectives. A second type of number marking in English is the inflection of verbs whose subject is third person singular (e.g. ‘I walk’ versus ‘She walks’. This regularity can also be used by English learners: If a verb is inflected with 's' they can infer that the subject is an individual, not a group.

As the above examples show, learners can therefore make inferences about the likely referents of novel labels by paying attention to the morpho-syntactic contexts in which those labels occur. This is known as (morpho)syntactic bootstrapping (Gleitman, 1990; Gleitman et al., 2005; Gleitman & Landau, 1994). The following section discusses evidence in support of children's use of this process.

1.3.2 Evidence in support of morpho-syntactic bootstrapping

Over the last 30 years especially, a substantial body of evidence has been developed showing that learners use regular mappings between form and meaning to make inferences about the meanings of novel labels. For example, Naigles (Naigles, 1990; Naigles & Kako, 1993) recognised that children may use the link between transitive syntax and causative events and between intransitive syntax and non-causative events to identify the meanings of novel verbs. As predicted, she found that 24-month-olds who heard the novel verbs in transitive frames (e.g. 'The duck is *glorping* the bunny!') associated them with causative actions (i.e. with an event in which a duck caused a rabbit to bend) while children who heard the novel verbs in intransitive frames (e.g. 'The duck and the bunny are *glorping*!') associated them with non-causative actions (i.e. with an event showing a duck and a rabbit bending independently). Two-year-old children are therefore able to use syntactic information (derived from either the syntactic frame itself, as Naigles (1990) argues, or from the number of verbal arguments, as Fisher (2002) suggests) to inform their interpretation of novel verbs. Evidence for 10-month-old infants' use of ditransitives to infer transfer of possession is provided by Gordon (2003). Evidence for 3- to 5-year-old children's use of clausal complements to infer mental states is provided by Papafragou, Cassidy, and Gleitman (2007). Both are discussed in some detail by Gleitman et al. (2005).

With regard to learners' understanding and use of adjectival morpho-syntax, there is evidence that children interpret labels introduced using adjectival syntax (e.g. 'This is a *zav* one!' 'This is a very *zavish* one!') as referring to properties (Hall, Waxman, & Hurwitz, 1993). In this study 2- and 4-year-olds were shown a target of a particular shape and material (e.g. a glass plate). This was then labelled with either count noun syntax ('This is a *zav*!') or adjectival syntax ('This is a *zav* one!'). They were then shown two more objects: a property match (e.g. a glass cup) and a shape match (e.g. a red plastic plate). Finally, they were asked to either find another *zav* (count noun condition) or find another *zav* one (adjective condition). The results showed that 4-year-olds tended to select the property match in the adjective condition, but not the count noun condition. The same was not true for 2-year-olds. This suggests that the ability to infer property meanings from adjectival syntax emerges between the ages of 2 and 4 years. This is supported by the fact that Taylor and Gelman (1988) found only a non-significant trend towards associating adjectival syntax with properties for 2-year-olds. Narrowing the developmental window down further, Smith, Jones, and Landau (1993) showed that 3-year-olds can make this inference. Follow-up work by Sorrentino (2001) shows that 3-year-olds do not blindly follow the morpho-syntactic context, but rather that their use of this as a cue interacts with their real-world knowledge. To illustrate, in sentences like 'This is *D/daxy*', the novel label is compatible with either a proper noun interpretation or an adjectival interpretation. Sorrentino (2001)'s results show that when the label is assigned to an animate object children interpret it as a proper name, but when it is assigned to an inanimate object they interpret is as an adjective (or perhaps a subordinate).

In addition to verbal and adjectival morpho-syntax, there is evidence that children are able to use nominal morpho-syntax to guide their interpretations of novel labels. One example of this is their ability to use both count noun syntax to infer that a novel label refers to a countable, individuated object, and to use mass noun syntax to infer that a novel label refers to a non-countable, non-individuated substance. In a very early study on bootstrapping, Brown (1957) showed that 3- and 4-year-old children interpreted a novel label as referring to a container if it was introduced using count noun syntax (e.g. 'Do you know what a *sib* is?') and as referring to the substance is inside the container if the label was introduced using mass noun syntax (e.g. 'Have you ever seen any *sib*?'). Similarly to English, French linguistically distinguishes count and mass nouns. A more recent study by Christophe, Milotte, Bernal, and Lidz (2008) found that 23-month-old French learners interpret labels as common nouns when they are introduced using common noun syntax. For children learning English, the onset of this type of behaviour may be slightly later, at around 30 months old (Soja, 1992). There is therefore converging evidence that young learners can use this kind of morpho-syntax to inform their understanding of novel nouns.

As discussed above, another relevant form-to-meaning mapping is the relationship between proper noun syntax and individual and common noun syntax and categories. There is evidence that children can use this information to make inferences about the likely meanings of novel nouns from the age of 24 months (Hall, Lee, & Belanger, 2001), but not before (Hall et al., 2001; Katz, Baker, & Macnamara, 1974). Jaswal and Markman (2001) provide supporting evidence of this with 2- and 3-year-old children.

There is also evidence that 4- and 5-year-olds can use the mismatch that occurs when a syntactically singular label is used to refer to a group of objects to learn the referents of collective nouns (Bloom & Keleman, 1996).

An additional pattern that learners can make use of is the relationship between singular number marking and individual objects, and plural number marking and groups of objects. This is of particular relevance to this thesis and it is therefore discussed in detail below.

1.3.3 Grammatical number marking

Many languages linguistically encode the semantic distinction between one and more than one. In English this encoding takes the form of grammatical number on the copula (i.e. the difference between ‘is’ and ‘are’), determiners, (e.g. the difference between ‘a’ and ‘some’ and ‘that’ and ‘those’) and nouns (e.g. the difference between uninflected singular nouns and plural nouns inflected with ‘s’). Children produce nouns marked as plural early in development and between the ages of 24 and 34 months produce the plural marker in 90% of all contexts in which it is required (Brown, 1973). However, despite this early aptitude, Berko (1958)’s classic study shows that children still have difficulty producing novel plurals by the age of 7. In terms of comprehension children understand grammatical number when it is marked on the determiner, noun and verb from the age of 24 months (Kouider, Halberda, Wood, & Carey, 2006) and when it is marked only on the noun from the age of 30 months (Jolly & Plunkett, 2008). In this way, in terms of both comprehension and production, children seem to acquire competence with grammatical number relatively early in development.

Proficiency with grammatical number is a useful skill because it allows learners to predict the referent of upcoming words: On hearing ‘is’ 2- and 3-year-olds tend to look at a singular referent rather than a plural referent and on hearing ‘are’ they tend to look at a plural referent rather than a singular referent (Lukyanenko, 2011; Lukyanenko & Fisher, 2016). In this way, the search space for candidate referents is limited by the preceding linguistic context. Following from this, grammatical number can also be used to disambiguate the referents of unknown words. For example, when 24-month-olds are introduced to novel labels in the context of singular morpho-syntax (e.g. ‘There is a *blicket*!’) they tend to look at an individual object rather than a group of objects (Kouider et al., 2006). Likewise, if they are introduced to a novel label marked as plural (e.g. ‘There are some *blickets*!’), they tend to look at a group of objects rather than an individual (Kouider et al., 2006). This is also the case for 30-, but not 24-month-olds, when number is marked only on the noun (Kouider et al., 2006). This suggests that children understand number marking on determiners and verbs before they understand number marking on nouns. I return to this point later.

Kouider et al. (2006)'s studies showed that although participants followed the number cue on both singular and plural trials they did so to a greater extent on plural trials than on singular trials (Davies, Xu Rattanasone, Schembri, & Demuth, 2019). One explanation for this is that children's understanding of plural number marking emerges before their understanding of singular number marking (Davies, Xu Rattanasone, & Demuth, 2019). This could be explained by the fact that in English singular number marking on nouns is null whereas plural number marking on nouns is explicit. One way to determine if this is the case would be to explore whether sensitivity to 'are' emerges before sensitivity to 'is'.

In addition to using grammatical number to disambiguate the referents of unknown words, children can also use grammatical number to learn the meaning of new labels (Jolly & Plunkett, 2008; Paquette-Smith & Johnson, 2016). Paquette-Smith and Johnson (2016) show this is the case from the age of 24 months. In their experiment participants were shown two creatures of one type and a single creature of a different type while hearing two plural sentences like 'These are nice *blickets*' and two singular sentences like 'Where is the *blicket*?'. This constituted a training phase. In the test phase, the referents were presented as either two pairs or two individuals and participants were asked to 'Look at the *blickets*' and 'Look at the *blicket*' respectively. In this phase, the number marking did not identify the intended referent and participants had to rely on associations formed during the training phase. That is, if Creature A were presented as an individual on singular trials and as a group on plural trials, they should look at Creature A when asked to look at the *blicket* or *blickets*. The results showed children did precisely this. Therefore, by the age of 24 months children can use number marking to inform their referential choices and learn the meanings of new words. In this experiment, number was variously marked on the verb, determiner and noun. A very similar study by Jolly and Plunkett (2008) found comparable results for 30-month-olds when number was marked only on the noun. This was not the case for 24-month-olds. In this way, there is converging evidence that children understand number marking on verbs and determiners before they understand number marking on nouns (see also Wood, Kouider, & Carey, 2009). These findings contradict the results of a recent study by Blossom (2013), in which 2-year-olds were shown to understand nominal number marking but not the difference between 'is' and 'are' One explanation for this is that this study considered the use of verbal morphology in the context of bare plurals (e.g. 'There are the sheep') and, as described above, children of this age expect agreement between subjects and verbs (Lukyanenko & Fisher, 2016). It is therefore possible that children did not follow the number marked verb in this context not because they did not understand the import of grammatical number, but because they read the bare plural as singular and therefore perceived a mismatch between the subject and the verb.

Section 1.2 above contained preliminary evidence that young children follow gestures that contain two elements more than they follow gestures comprising a single element. There is some evidence that a similar pattern can be observed with number marking. Specifically, Lukyanenko (2011) provides evidence that learners respond more to sentences that contain more number marked elements. To elaborate, this study shows that 3-year-old children who are able to use both verbal and nominal number marking respond more to sentences that include both types of number marking than to sentences that include only one type. In this study, participants were shown a picture of a single familiar object (e.g. an apple) and a picture of two other familiar objects (e.g. two cookies). At the same time, they heard sentences that either contained number marking on the noun and the verb (e.g. 'Where **are** the good cookie**s**?') or on the noun only ('Can you find the good cookie**s**?'). The results showed that participants looked at the group after hearing plural sentences and at the individual after hearing singular sentences in both conditions, but did so for longer after hearing the sentences with two number marked elements than after hearing sentences with one number marked element. This inspired my experiment exploring grammatical number marking. This experiment is discussed in more detail in the following section.

1.3.4 Introducing an experiment exploring children's understanding of grammatical number marking

The aim of this experiment is to explore Lukyanenko (2011)'s finding further and provide more evidence that children are more sensitive to doubly number marked sentences than to singly number marked sentences. Furthermore, it aims to determine whether this finding with familiar referents and conventional labels extends to unfamiliar referents and novel labels. As described above, in Lukyanenko (2011)'s study, participants were responding to the number marking on the copula, not (only) the number marking on the familiar noun (i.e. they started looking at the target after hearing the verb and before hearing the noun). Their behaviour cannot, therefore, be attributed to lexical knowledge of the nouns. Nonetheless, a task involving familiar labels and referents is inherently less demanding than it would be if it involved unfamiliar labels and referents. For this reason, behaviour in a task of the former kind does not necessarily predict performance in a task of the latter kind. In this thesis, this issue is addressed by comparing children's responses to sentences with number marking on the noun only versus sentences with number marking on the noun and verb in the context of unfamiliar labels and referents.

One limitation of previous studies exploring children's understanding and use of grammatical number is that the group of objects used as the target of plural number marking and the individual object serving as the target of the singular number marking belong to different categories (e.g. a group of lemon juicers and one honey spoon). By doing so, the task is made easier than the task children face in the real world. To illustrate, common nouns refer to whole categories and only objects of the same type can be labelled with the same basic-level category label: A group of cats and dogs cannot be referred to using 'cats' or 'dogs'. To follow grammatical number in the real world, learners must therefore be able to group together objects of like kind. It is not enough for a learner to look for groups after hearing a plural. Rather, she must look for a group of things of the same kind. The experiments using potential referents of different kinds limit the search space by providing one clearly defined group and one clearly defined individual. The task of delineating categories and working out 'what goes with what' is removed from the equation. To the best of my knowledge, all previous studies looking at children's use of grammatical number have used objects from different categories as the targets of the plural and singular labels. It is therefore possible that we have overestimated children's ability to follow and use grammatical number.

My experiment looking at grammatical number aims to address the limitation by using objects from the same category, (comparing for example one knife guard with a group of knife guards). I believe this task is more akin to the task children face in the real world and may therefore more accurately reflect 2- and 3-year-olds' ability to follow and use grammatical number in referential tasks. Note, however, that even with this modification the task is still referential: Participants are working out which potential referent a label is being used to draw attention to. However, because there is only one candidate, the participants do not have to work out which category of objects the label refers to. As such, it differs from the type of reference assignment task usually seen in studies of this kind (see for example Briganti & Cohen, 2011; Hollich et al., 2000; Houston-Price et al., 2006; Jolly & Plunkett, 2008; Kouider et al., 2006; Lukyanenko, 2011; Lukyanenko & Fisher, 2016; Paquette-Smith & Johnson, 2016; Pruden et al., 2006; Schafer & Plunkett, 1998). For clarity, this process will be referred to as understanding speaker reference (working out what speakers are using labels to draw attention to) rather than word reference (working out what category of object a novel label refers to).

This experiment therefore seeks to answer the following questions:

1. Are children more sensitive to the combination of nominal and verbal number marking than nominal number marking alone?
2. Do children follow grammatical number when the potential referents belong to the same category (e.g. one knife guard versus 14 knife guards)?

What might explain a preference for doubly number marked sentences over singly number marked sentences? As was the case with regard to the two element versus one element gestures, there are two possible explanations: the increased perceptual saliency of the doubly marked sentences, or their increased pragmatic intentionality. With regard to increased pragmatic intentionality, it is possible that if a speaker wishes to emphasise that she is referring to a group of objects she will use a sentence containing more number marked elements. This is discussed in more detail later. With regard to increased perceptual salience, it could be argued that singly number marked sentences are less audibly salient than doubly marked sentences. As was the case with the gestures, one way of teasing apart these explanations is to consider the issue developmentally. In particular, if the preference for doubly marked sentences is attributable to their greater intentionality, it is predicted that this preference will increase across development as children's ability to read referential intentions increases across development. It is not clear what, if any, developmental differences should be predicted if the preference for doubly marked sentences is due to their increased perceptual salience. As discussed above, evidence from word learning studies suggests that the role of the perceptual salience of the potential referents decreases across development. It is therefore possible that this is also the case regarding the salience of the cue itself. If so, if the preference for doubly marked sentences is due to their perceptual salience this preference may decrease across development. Alternatively, no developmental differences may be found. For these reasons I test children across a wide age range (2 to 3 years old), beginning at the point in development at which the relevant cues come in to play (24 months old). A second reason for testing this age range is to prepare for the final experiment presented here, which compares children's use of grammatical number and body orientation within this window. In this way, this experiment acts as a pre-test for Experiment 3.

These questions are returned to in Chapter 3, which presents an experiment designed to test them. The previous two sections presented evidence suggesting that presenting multiple cues of the same type (either grammatical or social) might encourage children to follow those cues. The following section considers how cues of different types can have a similar effect.

1.4 Solving the Reference Problem Using Multiple Cues

The following sections discuss how children use information of different types to solve the reference problem. The first section highlights some of these cues and details how children's use of them varies across the course of development. The second section considers the facilitative effect of multiple cues on children's visual attention. The third looks at evidence for and against the facilitative effect of multiple cues on children's ability to learn words. The final section introduces an experiment looking at children's use of social and grammatical cues in combination and in isolation.

1.4.1 Multiple cues and word learning

All theories of lexical acquisition acknowledge that children use many different types of information to learn the meanings of new words. Where they differ is in which types of cue they consider to be primary and which secondary. One theory that emphasises the role of multiple cues is the Emergentist Coalition Model (Golinkoff & Hirsh-Pasek, 2006; Hirsh-Pasek & Golinkoff, 2001, 2008; Hollich et al., 2000). On this account, children use perceptual, social and linguistic cues to learn words. The weights assigned to these cues vary across the course of development. Specifically, children weight strategies that they have learned to be more useful more heavily than those they have learned to be less so. As discussed previously, this process is known as Guided Distributional Learning. As a consequence of this learning mechanism, learners gradually shift from a reliance on perceptual cues (at around 10 months) to a reliance on social cues (at around 2 years). As noted above, some researchers have posited a second shift from reliance on social cues to a reliance on linguistic cues (Nappa et al., 2009; Paquette-Smith & Johnson, 2016). This account also acknowledges the role of specific word learning heuristics and suggests that these develop over time in response to children's previous word learning experiences. On this account, there are therefore multiple cues available to learners, although not all will be utilised by learners at all developmental stages. Given this, one might ask what happens when these cues interact. Is there a benefit associated with having multiple cues pointing to the same referent? The following section discusses the results of studies exploring this issue.

1.4.2 Following multiple cues

Hollich et al. (2000) investigate the roles of perceptual salience and gaze direction on young children's referential choices. In this experiment, participants were shown two unfamiliar objects: one interesting (e.g. a sparkly blue wand) and one less so (e.g. a white cabinet latch). The experimenter then looked at either the interesting object (coincident condition) or the less interesting object (conflict condition) and labelled it with an unfamiliar label ('Look! A *modi*!'). The results showed that 12-, 19- and 24-month-olds looked longer at the interesting object than the boring object in both conditions, but they did so for longer in the coincident condition than in the conflict condition. This suggests that the combination of cues (perceptual and social) were more effective in directing attention than the perceptual cue in isolation. This is comparable to Moore, Angelopoulos, and Bennet (1999)'s finding that 24-month-olds attended more to a target that was perceptually and socially cued than to a target that was only perceptually cued.

Paquette-Smith and Johnson (2016) provide evidence that this is also the case for social and grammatical cues. In their study, 24- and 25-month-olds were shown an individual unfamiliar animal and a pair of unfamiliar animals of a different kind. The speaker then produced a sentence with either singular number marking or plural number marking. In the convergent condition, the sentence was accompanied by the speaker looking towards the individual on singular trials and towards the pair on plural trials. In the divergent condition, the sentence was accompanied by the speaker looking towards the pair on singular trials and towards the individual on plural trials. In this way, multiple cues specified the target in the convergent conditions and only one specified the target in the divergent condition. The results showed that participants looked longer at the target in the convergent condition than in the divergent condition.[2]

Nappa et al. (2009) provide further evidence of the facilitative effect of multiple cues with regard to the combination of conceptual, social and linguistic cues. Their study explored 3-, 4-and 5-year-old children's acquisition of predicate perspective pairs such as 'chase' and 'flee' or 'give' and 'take'. Children of this age (and adults) have a conceptual bias towards interpreting ambiguous labels of this kind from the perspective of agent, i.e. 'chase' is preferred to 'flee', and 'give' is preferred to 'take', etc. This experiment explored the effect of social cues (i.e. gaze direction towards to the subject) and/or a linguistic cue (i.e. disambiguating sentences with full noun phrases) on this bias. The

[2] In this experiment the grammatically-cued referent was designated the target. It is not stated whether looks to the socially-cued referent were also higher in the convergent condition than the divergent condition.

results showed that participants looked longer at the target when it was signalled by conceptual, social and linguistic cues than when it was signalled by conceptual and social cues only. This provides further evidence that multiple cues facilitate children's attention to those cues.

Thus far, we have seen how young children follow multiple cues to a greater degree than single cues. This evidence comes from the training phases of word learning studies. Turning now to the test phases of those same studies, we see that multiple cues support children's referential decision making to a greater extent than single cues.

1.4.3 Using multiple cues

Returning to Hollich et al. (2000), recall that participants were shown two unfamiliar objects: one interesting (e.g. a sparkly blue wand) and one less so (e.g. a white cabinet latch). During the training phrase, the experimenter then looked at either the interesting object (coincident condition) or the less interesting object (conflict condition) and labelled it, e.g. 'Look! A *modi*!' During the test phase, the experimenter then moved out of sight and asked participants to 'Find the *modi*!'. The results showed that 19- and 24-month-olds infants looked at the target longer in both the coincident and conflict conditions. The 12-month-olds, however, did so only in the coincident condition, i.e. when perceptual and social cues were aligned. When they were out of alignment (i.e. the experimenter labelled the less interesting object), participants showed no evidence of associating the label with either object. In this way, although neither was sufficient for making referential decisions on their own, 12-month-olds were able to make referential decisions based on the two cues combined. This suggests that combining cues of two different types can facilitate children's uptake of those cues. In addition, these studies suggest that the role of perceptual salience in children's referential decision making decreases across development (contra Moore et al., 1999), while the role of gaze direction increases. This point is taken up again later.

Returning to the effect of multiple cues on word learning, the test phase of Paquette-Smith and Johnson (2016)'s experiment showed that when the target was signalled by gaze direction alone (Experiment 1), participants looked at the target more than the competitor by a difference of 56% to 44% (Paquette-Smith & Johnson, 2016, p. 332). When the target was signalled by gaze direction and number marking, however, (Experiment 2), the proportion of looks to target increased to 65%. The authors do not state whether this difference is significant, but the direction of the difference suggests the facilitative effect of multiple cues on word learning.

Similarly, the test phase of Nappa et al. (2009)'s experiment showed that 3-, 4- and 5-year-old children's referential decision making is facilitated by the presence of over-lapping conceptual, social and grammatical cues. Indeed, when these three cues were presented in alignment, participants selected the targeted interpretation of the novel label on approximately 80% of trials. When the social cue contradicted the conceptual and grammatical cue this was reduced to approximately 71% and when the grammatical cue contradicted the social and conceptual cue this was reduced to around 39% (calculations based on Nappa et al., 2009, p. 221, Figure 9). This suggests a positive correlation between multiple over-lapping cues and children's ability to learn words based on them. Furthermore, it suggests that by the age of 3 years learners weight grammatical cues more heavily than social cues. This point is returned to later.

Thus far, I have discussed how multiple overlapping cues can facilitate learners' uptake of those cues. However, this is not always the case, and some studies suggest that the presence of multiple cues does not have a facilitative effect. These are discussed in more detail below

In an early study looking at the children's use of social and perceptual cues, Moore et al. (1999) found that 24-month-olds did not look at a target for any longer when it was signalled by gaze direction and salience (i.e. in the match condition) than when it was signalled by gaze direction only (i.e. in the two control conditions). Similarly, Houston-Price et al. (2006) found no facilitative effect of the combination of social and statistical cues. In particular, 15-month-olds' attention to the target was no greater when it was signalled by both a statistical cue and gaze direction (Experiment 4) than when it was signalled by either gaze direction alone (Experiment 1) or the statistical cue alone (Experiment 2). This led the researchers to conclude that 'when two cues coincide to indicate the same referent, their utility is not simply additive in nature' (Houston-Price et al., 2006, p. 51).

In this section we have seen that the presence of multiple cues has a facilitative effect on children's ability to follow those cues (cue following) and also possibly on their ability to use these cues (cue use). In the following section we introduce an experiment exploring children's tendency to follow social and grammatical cues in combination and in isolation.

1.4.4 Introducing an experiment exploring children's understanding of grammatical and social cues

This experiment explores children's understanding of social and grammatical cues in contexts where the two cues point to the same referent (congruent condition) and in contexts where they point to different referents (incongruent condition). The prediction is that they will follow the cues more in the congruent condition than they do in either the incongruent condition or in either of the two previous experiments. It considers social and grammatical cues in particular because there is the possibility of interesting developmental differences in learners' tendency to follow these cues. To elaborate, as described above, developmental changes in children's use of perceptual cues and social cues led researchers to posit a learning mechanism known as 'Guided Distributional Learning' (for summaries see Hirsh-Pasek & Golinkoff, 2001, 2008). According to this account, the weights learners assign to different cues vary according to how useful those cues have been in the past. Consequently, children learn to weight social cues more heavily than perceptual cues because they are the more reliable guide to reference. Intuitively, this is the case: Boring objects have names as much as interesting objects do and although speakers will often label objects that they believe to be of interest to the child, this is not always the case (especially across development). As noted previously, Paquette-Smith and Johnson (2016) have suggested that a similar argument could be made with regard to children's use of social and grammatical cues (see also Nappa et al, 2009).

To reiterate, while a plural label will always and necessarily refer to a group of objects, a speaker will not necessarily be looking or pointing at the referent of a given label. Over the course of many naming episodes, speakers may therefore come to weight grammatical cues more heavily than social cues. Indeed, as discussed above, there is existing evidence that this is the case (i.e. Nappa et al., 2009). The following experiment has a congruent condition, where the social cue and the grammatical cue are in alignment, and an incongruent condition, where they are out of alignment. A developmental shift in children's weightings of social and grammatical cues may therefore be evidenced by their behaviour in the incongruent condition (with younger participants following the social cue and older participants following the grammatical cue).

This experiment therefore aims to answer the following questions:

1. Do children follow the combination of social and grammatical cues more than they follow either cue in isolation?
2. Does children's tendency to follow social versus grammatical cues change across development?

The experiment exploring this question is presented in Chapter 5.

All of the research questions posed here can be explored using a modified version of the Intermodal Preferential Looking Paradigm. In the following chapter I give an overview of this paradigm and discuss how it can be adapted for the purposes of this research.

Chapter 2 General Methodology

All the experiments presented in this thesis are based on the Intermodal Preferential Looking Paradigm (Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; for a recent review see Piotroski & Naigles, 2012). This paradigm was developed from the methodologies used by Huttenlocher, Smiley, and Charney (1983); Spelke (1979); and Thomas, Campos, Shucard, Ransay, and Shucard (1981). In its current form (see for example Paquette-Smith & Johnson, 2016), it closely resembles the concurrently developed looking-while-listening procedure (Fernald, Zangl, Portillo, & Marchman, 2008; Swingley, 2012).

The Intermodal Preferential Looking Paradigm (IPLP) is based on the assumption that if a listener hears a snippet of language (be it a word, phrase or sentence) they will tend to look at a stimulus that matches the language, rather than one which does not. In other words, it exploits the fact that people tend to look at things that are being spoken about (Rowland, 2014). In a typical preferential looking set-up, a child will be seated on their parent’s lap in front of two screens (to display the visual stimuli) with a speaker hidden in between them (to present the audio stimuli). The monitors will each display an image (for example a cat and a dog) and participants will hear a sentence referring to just one of them (e.g. ‘Look at the dog!’). If the participant has understood the sentence, they should orient their attention to the screen that matches the input (i.e. the screen showing the picture of the dog) rather than the screen that does not (i.e. the screen showing the picture of the cat). The display with the matching image is referred to as the ‘target’, and the display with the non-matching image as the ‘competitor’. To record their responses, a zoomed-in camera hidden in between the screens records participants as they view and hear the stimuli. These recordings are then coded frame-by-frame to indicate where the participants are looking at that point (typically whether they are looking to the display on the left, to the display on the right, or elsewhere). The dependent measure varies, but is often total time spent looking at the target display as a proportion of time spent looking at the target display and the competitor display combined. If the proportion of time spent looking at the target is significantly greater than the proportion of time spent looking at the competitor it is taken as evidence that participant has understood the sentence. In this way this method can be reliably used to assess children’s lexical comprehension (for the first example of this see Golinkoff et al., 1987).

Another prolific use of the IPLP is to test young children’s ability to learn new words. In the first demonstration of this, Schafer and Plunkett (1998) showed 12-month-old infants a picture of an unfamiliar object accompanied by a novel label (e.g. *bard*). Participants were then shown a second unfamiliar object accompanied by a second novel label (e.g. *sarl*). This was repeated six times for each pair. Finally, participants were shown pictures of both the unfamiliar objects accompanied by either one of the previously heard novel labels or a third novel label (*geek*). The results showed the participants looked longer at the picture previously labelled *bard* when they heard *bard* and longer at the picture previously labelled *sarl* when the heard *sarl.* In both cases they did so for longer than when they heard *geek* (albeit non-significantly). This suggests that after six repetitions, 15-month-olds can learn to associate two novel labels with specific unfamiliar objects. In this way, the IPLP can be used to assess children's ability to learn new lexical items.

Beyond this, the IPLP can also be used to explore young children's understanding of the meanings encoded by different types of morphology and syntax. For example, English linguistically encodes the distinction between 'one' and 'more than one' such that groups of objects are referred to using plural number marking and single objects are referred to using singular number marking. As discussed previously, several studies show that young children can use this feature of English morpho-syntax to direct their attention and inform their referential choices. Many of these studies used the IPLP (e.g. Jolly & Plunkett, 2008; Kouider et al., 2006; Paquette-Smith & Johnson, 2016). To illustrate the use of the IPLP to direct visual attention I present the methodology employed by Kouider et al. (2006) in more detail. To my knowledge, this is the first study that uses the IPLP to explore children's ability to follow (but not use) cues for lexical acquisition.

In this study, 24-month-old learners were seated in front of two screens. One showed a picture of a group of unusual objects (e.g. eight honey-spoons) and the other a single unusual object belonging to a different category (e.g. a single lemon-juicer). Participants then heard sentences containing plural number marking, like 'There **are some** *blicket***s**!' and sentences containing singular number marking, like 'There**'s a** *blicket*!'. The rationale is that if the participants understand the semantic import of grammatical number they should look longer at the group of objects after hearing plural sentences and longer at the individual object after hearing singular sentences. This was precisely what the participants did. As such, the IPLP can be used to demonstrate young children's understanding of the relationship between morpho-syntax and meaning.

In the examples given above, the experiments show that learners behaviour changes depending on the types of potential referents they are presented with and depending on the linguistic context labels are presented in. Another relevant factor in learners' behaviour is the way speakers interact with potential referents. The IPLP can be modified to incorporate this type of information. Specifically, in the Interactive Intermodal Preferential Looking Paradigm (IIPLP), rather than being pre-recorded the sentences are produced 'live' by one of the experimenters who is able to interact with the potential referents. For example, in a series of preferential-looking studies exploring the role of social eye-gaze in young learners' referential choices, Hollich et al. (2000) showed infants two unfamiliar objects. Rather than displaying the pictures on a screen, however, the experimenter stood across a table from the participant in front of a board to which the potential referents were attached (one on left of the board and the other on the right). During each trial, the experimenter looked at one of the objects and labelled it (e.g. 'Look! A *modi*!') and then moved out of sight. Participants were then asked to 'Find the *modi*!'. Twenty-four-month-olds looked longer at the object the speaker looked at than the one she ignored. In this way, the IPLP can be used to assess the effect of speaker's non-verbal behaviour on learners' interpretation of novel labels.

In this experiment the speaker was 'live' and the potential referents were all physical objects. More recently, IIPLP studies have also successfully been run with a video-recorded speaker interacting with images of potential referents. (see for example Briganti & Cohen, 2011; Houston-Price et al., 2006; Nappa et al., 2009; Paquette-Smith & Johnson, 2016). One advantage of this method is that the behaviour of the speaker is identical across trials and across participants and there is no possibility for the speaker to inadvertently cue the participant. A disadvantage is that participants may be aware that the images of potential referents superimposed on the screen are not really there and cannot be seen by the actors. That is, participants may not be willing to suspend their disbelief. Nonetheless, this method has successfully been used with children up to the age of 5 years (Nappa et al., 2009) and will be adopted for use with the 2- and 3-year-olds in the present studies.

As noted above, the IPLP in all its forms is based on the assumption that both children and adults look at things that are being spoken about. Finding that a participant looks longer at Object A than Object B after hearing 'Find the *zoot'* is therefore taken as evidence that they believe '*zoot'* refers to Object A. This assumption has been criticised, however, on the basis that an increase in visual attention towards an object does not necessarily mean a referential decision has been made (see for example Jolly & Plunkett, 2008; Paquette-Smith & Johnson, 2016; Pruden et al., 2006). In light of this, some preferential looking studies include more stringent measures of referential understanding.

One way of assessing referential understanding more rigorously is to separate the experiment into a training phase and test phase. For example, in their training phase Paquette-Smith and Johnson (2016) showed 24-month-old participants a group of two objects and an individual objects in the context of sentences with singular number marking or plural number marking. For example, participants might have heard ‘These are nice *zurpels*’ and seen a single green creature and two purple creatures. In the test phase the pictures and sentences were matched in number: Participants either saw two green creatures and two purple creatures and heard plural sentences like ‘Can you see the *zurpels*?’ or saw a single purple creature and a single green creature and heard singular sentences like ‘Can you see the *zurpel*?’. If they understood the number marking in the training phase, they should look longer at the pair or the individual that was previously picked out by the number marking. The results showed this was the case for the 24-month-old participants. In this experiment, participants looked longer at the target in both the training phase and the test phase. This suggests that increased visual attention to the target was indeed indicative of referential understanding.

Pruden et al. (2006) use an even more stringent test of referential understanding. This study explored the role of perceptual saliency on children’s referential decisions. In this experiment, 10-month-old learners were shown two unfamiliar objects: one interesting (e.g. a sparkly blue wand) and one less so (e.g. a white cabinet latch) in the context of a novel label (e.g. ‘Look! A *modi*!’). As predicted, infants looked longer at the more interesting object during the training phase and then selected the interesting object when asked to ‘Find the *modi*’ during the test phase. As a follow-up test of referential understanding, the experimenters then introduced a second label (e.g. ‘Look! A *glorp*!’). As noted above, from the age of 17 months children have a reliable tendency to avoid having two labels for one object. The rationale for introducing a second novel label was therefore that if infants associated the first label with the interesting object they should avoid assigning it a second label and instead associate the newly-heard novel label with the boring object. This was borne out: Participants increased looking to the boring object on hearing the second novel label, and then looked back to the interesting object on hearing the first novel label. In this study, in both conditions participants looked longer at the interesting object than the boring object in the training phase. The was also the case in the test phase. This therefore provides more evidence that participants behaviour in the training phase predict later performance in the test phase and suggests that in the context of preferential looking, increased visual attention reliably correlates with referential understanding. For this reason, the experiments presented here use looking-time as a proxy for referential understanding and do not include secondary or tertiary tests. As such, the experiments presented here explore the role of cue quantity on children's ability to follow cues, but not on their ability to use those cues. I suspect, however, that increased visual attention to a target during training will predict better learning and retention of that label at test. In this way, these experiments represent a first step in assessing the role of cue quantity in word learning. This point is picked up again in Chapter 6.

The above examples show how the IPLP can be used to assess children's use of both grammatical cues and gestural cues in establishing reference. Using the IPLP therefore allows me to use the same methodology to address all my research questions. The experiments presented here use a modified version of the IPLP in which two types of potential referent (groups of objects and a single object) are presented against a video-recording of a speaker who draws attention to them using sentences containing either nominal number marking or nominal and verbal number marking (Experiment 1), or using gestures encompassing either body orientation or body orientation and pointing (Experiment 2). Experiment 3 looks at the combination of these cues and considers body orientation and nominal number marking. Experiment 1 is presented in Chapter 3, Experiment 2 in Chapter 4 and Experiment 3 in Chapter 5.

Chapter 3 Experiment 1: Grammatical Number

3.1 Aims and Predictions

Section 1.3 discussed how children follow grammatical number from the age of 24 months when it is marked on verbs (Kouider et al., 2006), and from the age of 30 months when it is marked on nouns (Jolly & Plunkett, 2008). It also presented preliminary evidence that children who are able to follow both types of number marking (i.e. 3-year-olds) are more sensitive to doubly marked sentences (i.e. sentences that contain number marking on the noun and the verb) than singly marked sentences (i.e. sentences that contain noun only). This was found in an experiment involving familiar objects and labels (Lukyanenko, 2011). One of the aims of the present experiment is to bring more evidence to bear on this subject and explore whether this finding extends to unfamiliar objects and labels. As discussed, there are two possible ways of explaining an observed preference for the doubly marked sentences: their greater perceptual salience or their greater pragmatic intentionality. One way of distinguishing between these explanations is to look for developmental differences: If a preference for the doubly marked sentences is attributable to their greater pragmatic intentionality, the tendency should increase across development, but if it is attributable to their increased perceptual salience it should not (see Section 1.2.2 for details). For this reason, the experiments tests children across a wide age range, namely 2- and 3-year-olds. Based on the studies described above it is predicted that the youngest participants will understand the verbal number marking, but not the nominal number marking and will therefore follow the doubly marked sentences but not the singly marked sentences. For older learners, it is predicted that they will attend to both doubly and singly marked sentences, but more so to the former than the latter. Within the group of participants who attend to both cues, I look for further developmental differences in order to elucidate the origin of any observed preference for nominal and verbal number marking as described above.

Section 1.3 also described how all previous studies concerning children’s understanding of grammatical number used potential referents belonging to different categories (e.g. one honey spoon versus multiple lemon juicers). I argued that this makes the task easier for participants by eliminating the categorisation component (the question of what goes with what) and that previous studies may therefore overestimate children’s ability to follow grammatical number. This study aims to remedy this by exploring 2- and 3-year-old children’s ability to follow grammatical number when the potential referents belong to the same category (e.g. one knife guard versus multiple knife guards).

My first research question is therefore as follows:

RQ1: Do 2- and 3-year-olds' follow grammatical number when the potential referents belong to the same category (e.g. one knife guard versus 14 knife guards)? If so, do they do so more on hearing nominal and verbal number marking than nominal number marking alone? Does this change across development?

If participants are sensitive to the number marking they should pay more attention to the group after hearing plural sentences and to the individual after hearing singular sentences. If they are more sensitive to nominal and verbal marking than nominal marking they should do so more in Condition 2 than Condition 1. The main independent variables are therefore number and condition. As described above, age is a third variable of interest.

A final variable of interest is allomorph. Specifically, there is some evidence that 2- and 3-year-olds old children pay more attention to plural labels that are inflected with the /s/ allomorph than labels inflected /z/ (Davies, Xu Rattanasone, & Demuth, 2016). The explanation given for this is that the longer duration of /s/ relative to /z/ makes it more salient and easier to segment from the speech stream (Davies et al., 2016). For this reason it is predicted that participants will show more sensitivity to the labels inflected with the /s/ allomorph than those inflected with the /z/ allomorph.

3.2 Methodology

3.2.1 Participants

Thirty-three 2- (n = 15, 9 girls) and 3-year-olds (n = 18, 12 girls) participated in this experiment (*M* = 3;0, Range = 1;11-3;10. All participants came from monolingual English-speaking homes (n = 32) or homes where English was the dominant language (n = 1). Participants were primarily recruited through local online parenting forums and university mailing lists. Parents were given a verbal and written description of the experiment (see Appendix 1) before giving informed consent on behalf of their children (see Appendix 2). As compensation for their time, parents were given a £2.50 voucher for a local coffee shop and children were given a sticker book. The experiment was approved by the University of Edinburgh Philosophy, Psychology and Language Sciences Ethics Committee. Seventeen additional children (seven 2-year-olds and ten 3-year-olds) participated in the experiment, but were excluded from analysis for the following reasons: restlessness (n = 12), equipment failure (n = 1) and trial loss (n = 4, see section 3.3 for details).

3.2.2 Design

The experiment used a 2x2 factorial design. The first variable was condition: number making on the noun (Condition 1) or number marking on the noun and verb (Condition 2). The second variable was number (singular or plural). These variables were fully crossed and varied within subjects. The plural allomorph taken by nominal labels is balanced within conditions for novel labels and fixed for familiar labels.

The objects displayed were of the same type (e.g. knife guards) and different number and colours (e.g. 14 purple knife guards and a single yellow knife guard). As shown in Figure 1 each object was presented in two different colours: yellow and purple or green and red for unfamiliar objects and blue and green for familiar objects.

Participants completed two different types of trial: experimental trials in which unfamiliar objects were presented with novel labels (e.g. 'There are the *zoots!'*) and control trials in which familiar objects were presented with their conventional labels (e.g. 'There are the spoons!'). The control trials were included to break up the task and allow a secondary analysis if participants do not behave as expected on the experimental trials.

Each participant completed 20 trials: 16 experimental trials (involving unfamiliar objects and unfamiliar labels) and 4 control trials (involving familiar objects and familiar labels). For the experimental trials, 4 novel objects and 4 unfamiliar labels were used. Each object/label pairing occurred twice in each condition (once singular and once plural). For control trials, two familiar objects and two familiar labels were used and each occurred once per condition (either as singular or as plural). Four versions of the experiment were created. Two of these varied the pairing between the unfamiliar objects and the unfamiliar labels[3] and the other two versions varied the position of the target. To illustrate, in Version 1 (List A) Object 1 was paired with Label 1 and the target appeared on the left; in Version 2 (List B) Object 1 was again paired with Label 1, but this time the target appeared on the right; in Version 3 (List C) Object 1 was paired with Label 3 and the target appeared on the left and in Version 4 (List D) Object 1 was paired again paired with Label 3, but this time the target appeared on the right. Within each of these lists the target was one colour on half the trials and a different colour on the other half. This was balanced within trial type such that for experimental trials the target was red four times, green four times, purple four times and yellow four times and for control trials it was green twice and blue twice. A final factor was whether the actor turned clockwise

[3] The object-label pairing was varied for unfamiliar objects and labels only; familiar objects were always presented with their conventional labels.

or anti-clockwise. This was yoked to labels such that half of the labels always occurred with a clockwise turn and half always occurred with an anti-clockwise turn. The lists are reproduced in Appendix 3. Participants were assigned to lists pseudo-randomly such that equal numbers of participants completed each one.[4]

3.2.3 Materials

Audio stimuli

Twenty-four sentences served as audio stimuli. Twelve of these consisted of the carrier phrase ‘I found...’ followed by a noun phrase (e.g. ‘the *zoots*’). These served as the stimuli for Condition 1. The remainder consisted of either the singular carrier phrase ‘There’s...’ or a plural carrier phrase ‘There’re...’ both followed by a noun phrase. These served as the stimuli for Condition 2. As shown in Table 1, in both conditions the noun phrase was ‘the X’, where X was either a real English word (n=2) or a nonsense word (n=4). Each label occurred in the singular and the plural and in each sentence frame, i.e. ‘I found the X’, ‘I found the Xs’, ‘There’s the X’ and ‘There’re the Xs’.

Table 1: Labels used in Experiments 1-3

Novel Labels			Familiar Label		
Singular	**Plural**	**Allomorph**	**Singular**	**Plural**	**Allomorph**
Feek (/fik/)	*Feeks* (/fiks/)	-s	Chair	Chairs	-z
Zoot (/zut/)	*Zoots* (/zuts/)	-s	Spoon	Spoons	-z
Glaive (/glev/)	*Glaives* (/glevz/)	-z			
Tulb (/tʌlb/)	*Tulbs* (/tʌlbz/)	-z			

The real English words were ‘chair’ and ‘spoon’. These were selected as being highly familiar to even the youngest participants: According to the Crosslingusitic Lexical Norms database (Dale & Fenson, 1996), 90.9% of 18-month-olds know the word ‘chair’ and 92.4% of 18-months-olds know the word ‘spoon’. The youngest children tested here were 24 months old. All nonsense words were consistent with native English phonotactics and had monosyllabic roots. When marked as plural, two labels took the /s/ allomorph and two took the /z/ allomorph. The labels were designed to be phonologically

[4] Following participant loss, the final proportions were as follows: 8/33 completed List A; 7/33 completed List B; 8/33 completed List C; and 10/33 completed List D.

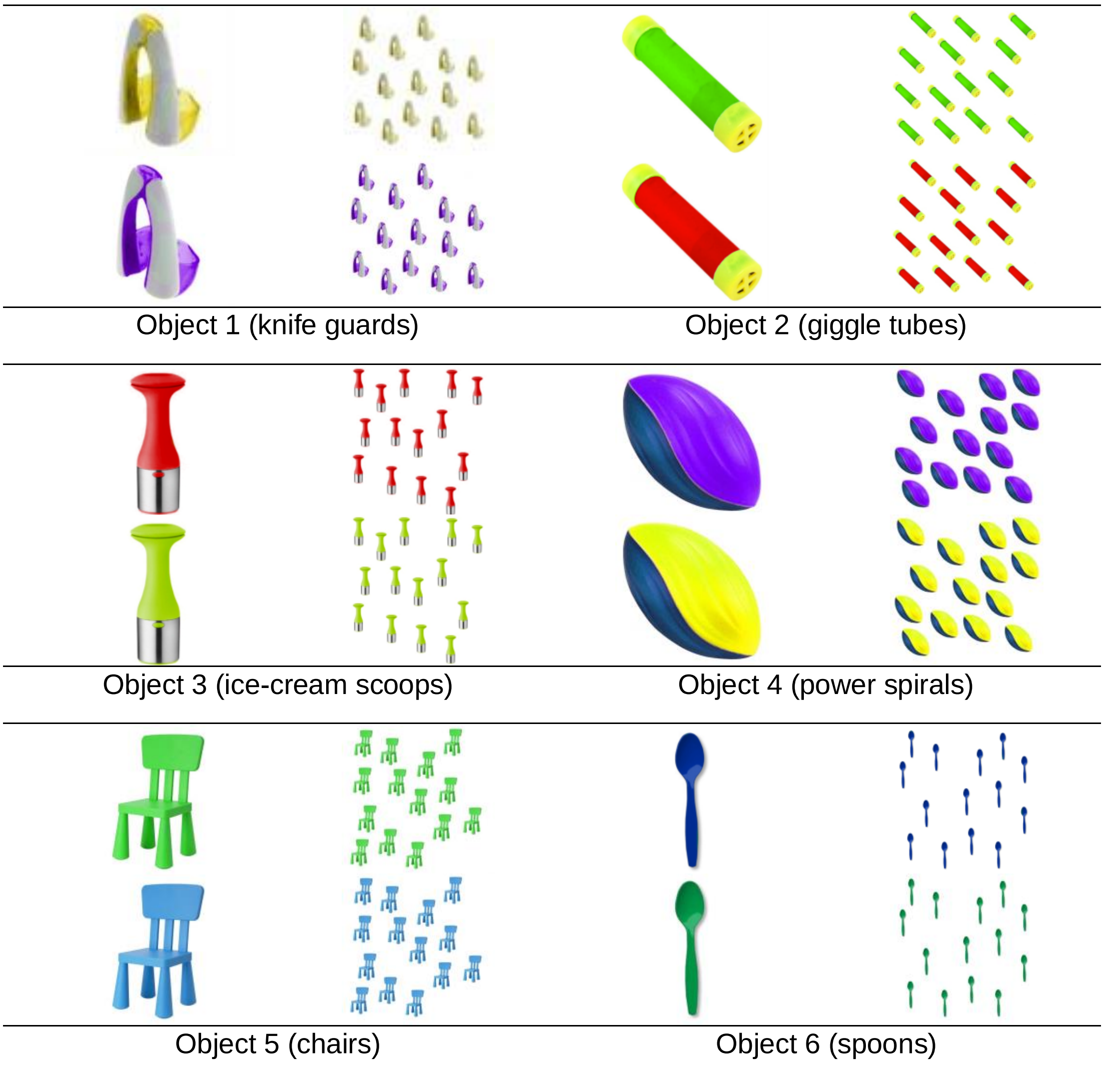
distinct from each other: Each had a different consonantal onset and coda and contained a different vowel.

The sentences were initially recorded in their entirety (e.g. ‘I found the *zoots*!’). A single example of the carrier phrase and a singular and plural version of each of the noun phrases (e.g. ‘the *zoot’* or ‘the *zoots’*) were then isolated from the originals and spliced together to create the experimental stimuli. All audio-stimuli were produced by a female native-English-speaker with a North American accent using child directed speech. This actor was chosen for her child-friendly voice and demeanour. Given the prevalence of American children’s television programmes and films (e.g. most Disney and Pixar films and programmes such as Blaze and the Monster Machines, Dora the Explorer, PJ Masks, Paw Patrol), her accent is likely to be highly familiar to children growing up in Scotland.

Visual stimuli

Digital photographs of six objects served as visual stimuli. Of these objects, four were unfamiliar (i.e. participants were not expected to know their names) and two were familiar (i.e. chairs and spoons). All objects were primarily one colour and did not have obvious subparts. Each object was presented as a single large image and as a set of 14 smaller images. These displays were designed so that the total surface area of the large image and the combined surface area of the smaller images were approximately equal. A colour-change tool was used to create a second version of each display in a different colour. All the images serving as visual stimuli are reproduced in Figure 1.

Figure 1: Images used in Experiments 1-3



3.2.4 Trial procedure

In each trial two images were displayed (see Figure 2). These images were always a single object and a group of objects presented against a white rectangular background. The rectangles were positioned in the upper left and upper right corners of the screen such that the outside edge of each rectangle was 1.5cm from the edge of the screen, 2cm from the top of the screen and the inside edges of the rectangles were 31cm apart. The rectangles measured 22cm by 24cm. These images were then embedded in videos. In these videos, the images were super-imposed above a female actor standing in front of a wooden slatted wall such that the images appeared to be on the wall with the actor standing between them (see Figure 2). As the trial progressed the actor moved and verbally interacted with the pictures. The scene measured 77cm by 62cm and was presented on a wide screen monitor.

Figure 2: Visual stimuli in situ, Experiment 1

At the start of the trial both rectangles were empty and the actor stood in between them facing forward. After 1000ms a ‘boing’ sounds was played and the images appeared in the blank rectangles. ‘Hearing’ this, the actor gasped, said ‘What’s that’ and turned to face the images on the wall behind her. Apparently seeing them, she said ‘Look!’ followed 1720ms later by a singular or plural sentence from either Condition 1 or Condition 2. She then continued facing the wall until the ‘boing’ sound was played again and the images disappeared. She then turned to face the camera and the trial ended. The entire trial lasted 16 seconds. Further details of within-trial timings can be found in Figure 3.[5]

[5] As shown in 3, the timings varied across condition. This is because the dividing point for the pre- and post-stimulus windows was set at the offset of ‘found’ in Condition 1 (e.g. at | in I found | the *zoot*!’ ‘I found | the *zoots*!’ and at the offset of ‘there’ in Condition 2 (e.g. at | in ‘There | ‘s the *zoot*!’ and ‘There | ‘re the *zoots*!’/). These points were not anchored to the same point in the trial and we originally envisaged using a different dividing point. As such, the location of the pre- and post-stimulus windows varies across condition and number. More details are given in Section 3.4.2. The locations of the windows were chosen the ensure they were comparable across condition and number, i.e. by containing neutral sentential information. A second difference is the location of ‘look’, at 4560ms in Condition 1 and at 5280ms in Condition 2. In both cases this is 1720ms before the onset of the sentence.

Figure 3: Trial progression, Experiment 1

Time (ms)	Audio		Visual		
	Condition 1: nominal only	**Condition 2: nominal and verbal**	**Actor**	**Object**	
0			Facing camera	Absent	
1000	Boing!	Boing!’	Facing camera	Present	
2080	Gasp!	Gasp!	Facing camera	Present	
3160	‘Let’s see!’	‘What’s that?’	Turning to wall	Present	
4000			Turning to wall	Present	
5560	‘Look!’		Facing wall	Present	
6000			Facing wall	Present	
6280		‘Look!’	Facing wall	Present	
7280	‘I found...		Facing wall	Present	
8000	... the *zoot/s*!’	‘There’s/’re the *zoot/s*!’	Facing wall	Present	
9000			Facing wall	Present	
10000			Facing wall	Present	
11000			Facing wall	Present	
12000			Facing wall	Present	
13000	Boing!	Boing!	Turning to camera	Absent	
14000			Turning to camera	Absent	
15000			Turning to camera	Absent	
16000			Facing camera	Absent	

3.2.5 Session procedure

Participants were seated on their parent’s lap approximately 150cm in front of a wide-screen monitor. The visual stimuli were displayed on the monitor and the audio stimuli were played through internal speakers. Participants’ eye-movements were recorded at a rate of 25 frames per second using a remote-controlled video camera hidden under the monitor. To avoid cueing their children, parents were asked to listen to music through headphones and to either close their eyes or avoid looking at the screen throughout. Parents were assured that if their child became restless they could reassure them and encourage them to keep watching, but were asked not to interact with them otherwise. Finally, parents were reminded that they could stop the experiment at any time.

The session began with the screen displaying bubbles moving to the sound of children laughing. This was used throughout the session as an attention-getter and was played between trials until participants (re)focussed on the screen. At the start of the experiment, the attention-getter was followed by a short animation in which a ladybird bounced around the screen. This animation was included to give the participants time to get used to the environment and to familiarize themselves with the dimensions of the screen. This animation was followed by the attention-getter and then the first trial. The first trial, and all subsequent trials, were initiated by the experimenter when the participant was judged to be paying attention to centre of the screen (i.e. looking straight ahead). If a participant became distracted during a trial that trial continued until the video was played out and was followed by the attention-getter as usual. The next trial was not initiated until the participant focussed on the attention-getter.

All participants completed 20 trials: 16 experimental (unfamiliar objects with novel labels) and 4 control (familiar objects with conventional labels). Trials were presented in blocks of five. Within each block there were four experimental trials and one control trial. The trial order was fixed such that participants always completed two experimental trials followed by one control trial and then two more experimental trials. There were no other conditions on the order of presentation. This order was imposed so that the control trials were evenly distributed across the course of the experiment and participants had regular breaks from experimental trials. As additional breaks, the blocks were separated by four 16s animated sequences in which a cake, an umbrella, a ball or a balloon moved around on a brightly coloured background. The order and presentation of the stimuli were controlled using Habit X (Cohen, Atkinson, & Chaput, 2004) running on a Macintosh computer.

3.3 Coding

The video recordings showed the face of the participant and an insert in the top left corner showed what the participant was watching on the screen. As is standard, videos were coded ‘blind’ so that the coder did not know which condition a given trial belonged to. This was achieved by obscuring the actor in the insert and leaving the displays visible. Videos were coded frame-by-frame using four coding categories: ‘left’ (used when the participant was looking at the display on the left), ‘right’ (used when the participant was looking at the display on the right), ‘speaker’ (used when the participant was looking at the actor) and ‘elsewhere’ (used when the participant was looking anywhere else).[6] If a participant blinked for 5 frames or fewer and continued looking in the same direction, the frames were coded as if the blink had not happened. If a participant blinked for 5 frames or fewer and changed looking direction during the blink, the frames on which she was blinking were coded 'elsewhere'. If the blink lasted more than 5 frames the frames were coded 'elsewhere'. Blinks were defined as frames where participants either had their eyes fully closed or closed to an extent that the direction of gaze could not be determined. Videos were coded from when the visual stimuli appeared on the screen to when they disappeared.

A second coder independently coded complete data sets from eight participants (22% of the useable data). Four were randomly selected from the 2-year-old participants and four from the 3-year-old participants. Agreement between the coders was high. When all four coding categories (left, right, speaker and elsewhere) were used the mean agreement between coders was 91% (SD = 3.56, Cohen's κ = 0.87). When the 'speaker' and 'elsewhere' codes were collapsed agreement increased to 94% (SD = 2.25, Cohen's κ = 0.9). These scores indicate a 'near perfect' degree of agreement (Landis & Koch, 1977). When the coders disagreed, the first coder's decision was used. Videos were coded such that coders did not know which experimental condition any given trial belonged to. The data were then filtered such that any trial on which greater than or equal to 20% of frames[7] were coded 'elsewhere' was excluded from analysis.[8]

[6] In this experiment, one participant spontaneously pointed at the left display. In doing so, her arm obscured or partially obscured her eyes for approximately 20 frames. On these occasions, her direction of point, rather than her direction of gaze was recorded.

[7] I settled on a criterion of 20% as, across the three experiments, this dovetailed with my impressions of which participants should be excluded due to restlessness. That is, when the criterion was set at 20% only participants who had been independently judged as being not restless satisfied the additional criterion that all participants contribute at least one trial per condition; when a criterion of greater than 20% was used restless participants became eligible. Likewise, when a criterion of less than 20% was used, non-restless participants became ineligible. In this way, the criterion of 20% supported my independent judgements about which participants were sufficiently engaged in the task to be included in the analysis. Regarding analysis of this data, this criterion (and those described below) reduces the number of data points entered into the analysis. This then reduces the power of the experiment and increases the possibility of Type 2 errors. However, it also reduces the amount of noise in that data and in doing so reduces the possibility of Type 1 errors. To investigate whether this data loss affected the results, I also ran the analysis with these data included. Comparable results were obtained.

[8] Two hundred and two out of 740 trials were excluded in this way. This equates to approximately 27% of the useable data collected. Following this loss, 4 participants failed to contribute at least one trial per condition. As reported in 3.2.1, their data (33 trials) were excluded from further analysis. One additional trial was lost due to equipment failure so the total number of trials available for analysis was 504. Of these, 402 were experimental trials and 102 were control trials. The following analyses are conducted with the experimental trials only.

3.4 Analysis and Results

3.4.1 Eye movements

The analysis begins by considering participants' gaze-direction data to four visual areas: target, competitor, speaker and elsewhere. Overall, the data suggest participants were on-task and following the events on the screen. As illustrated in Figures 10-13 below, there is a peak in attention to the displays when the stimuli first appear on screen (1000ms) with an initial preference for the individual object. This early attention to the displays is followed by a period of sustained attention to the speaker as she 'hears' the objects appear and turns to look at them (c.2250-4750ms). Beyond this, looking behaviour varies according to condition and number. In Condition 1 there is a second peak in looks-to-speaker after she begin speaking. This is not the case in Condition 2. This possibly occurs because the first person pronoun used in the sentences in Condition 1 (i.e. "I found the…") directs attention to the speaker, whereas the dummy subject 'there' used in the sentences in Condition 2 (i.e. "There's/There're the…") directs attention away from the speaker and on to a third referent. With the exception of plural trials in Condition 2 (Figure 7) there is a final peak in looks-to-speaker in the post-stimulus analysis window. This suggests that rather than looking at the appropriate display after hearing the stimuli, participants tended to look at the speaker. One explanation for this is that they are waiting for the speaker to provide more disambiguating information. For plural trials in Condition 2, however, there is no final peak in looks-to-speaker, and looks-to-group are consistently higher than looks-to-individual. This suggests participants may be sensitive to the plural stimuli in Condition 2. The following sections explore whether these first impressions are borne out by statistical analysis.

Figure 4: Looks to group (red), individual (green), speaker (blue) and elsewhere (purple), Experiment 1, Condition 1, singular trials

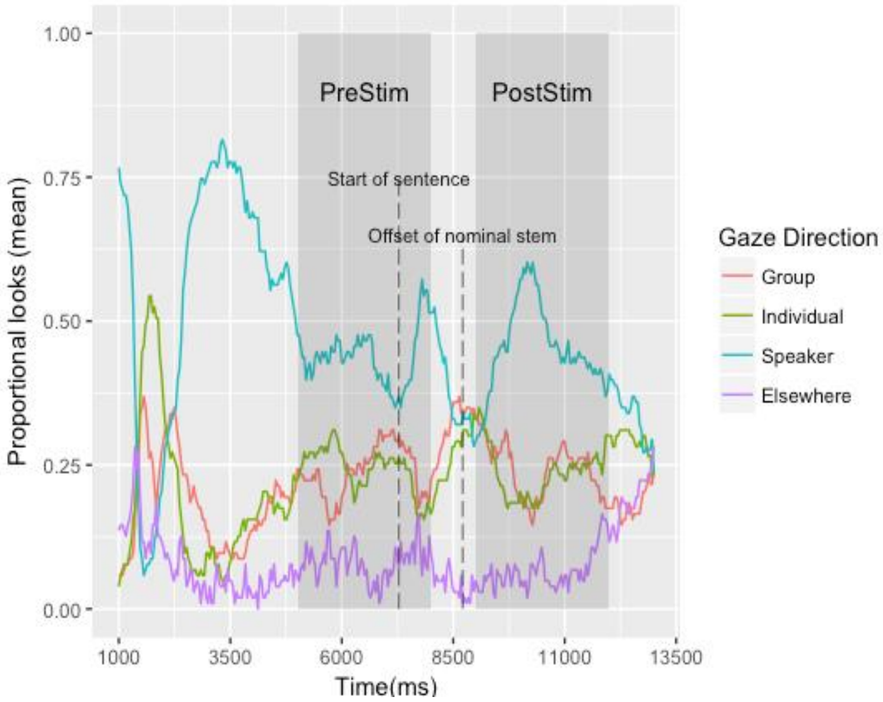


Figure 5: Looks to group (red), individual (green), speaker (blue) and elsewhere (purple), Experiment 1, Condition 1, plural trials

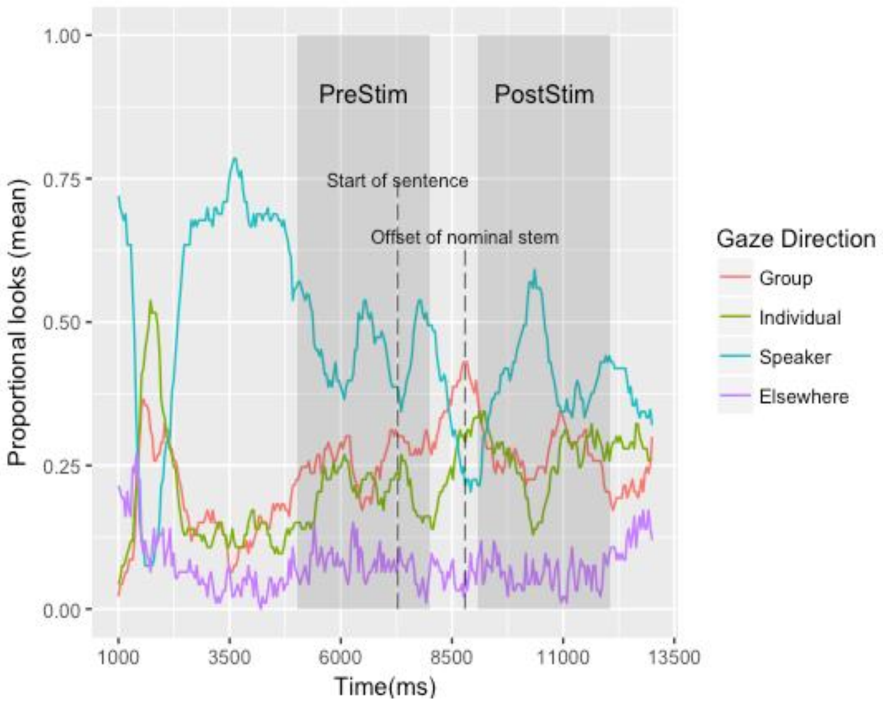


Figure 6: Looks to group (red), individual (green), speaker (blue) and elsewhere (purple), Experiment 1, Condition 2, singular trials

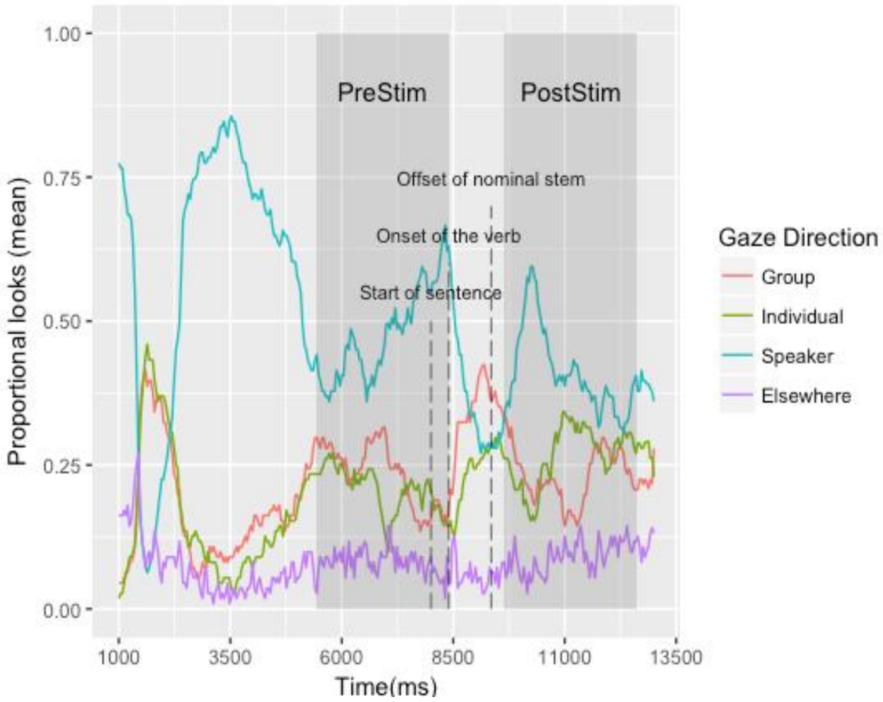


Figure 7: Looks to group (red), individual (green), speaker (blue) and elsewhere (purple), Experiment 1, Condition 2, plural trials

3.4.2 Proportional looks

Analysis

The dependent variable in this analysis is proportion of looks-to-group. This is similar to the proportion of looks-to-target measure used by Kouider et al. (2006) and Paquette-Smith and Johnson (2016). Proportion of looks-to-group is calculated by subtracting the pre-stimulus proportion of looks-to-group from the post-stimulus proportion of looks-to-group to give a difference score.[9] A positive difference score indicates a post-stimulus increase in looks-to-group and a negative difference score indicates a post-stimulus increase in looks-to-individual. If participants respond appropriately to the singular and plural stimuli there should be a post-stimulus increase in looks-to-group on plural trials and a post-stimulus decrease in looks-to-group on singular trials. If this is the case, difference scores should be significantly higher than zero on plural trials and significantly lower than zero on singular trials. The duration of the analysis windows is set at 2980ms. This is the longest window possible given the configuration of stimuli[10] and is broadly in line with a study using a very similar methodology in which the duration was set at 3500ms (Kouider et al., 2006).[11] In Condition 1, the pre-stimulus window extended back from the offset of ‘found’, e.g. | in ‘I found | the *zoot*!’ (singular) and ‘I found | the *zoots*’ (plural). In Condition 2, it extended back from the offset of ‘there’, e.g. | in ‘There | ‘s the *zoot*!’ (singular) and ‘There | ‘re the *zoots*!’ (plural). In Condition 1 this point is located at 7000ms and the pre-stimulus window runs from 5020ms to 8000ms. In Condition 2, this point is located at 8400 for singular trials and 8320 for plural trials. The pre-stimulus window therefore runs from 5420 to 8400 for singular trials and from 5340

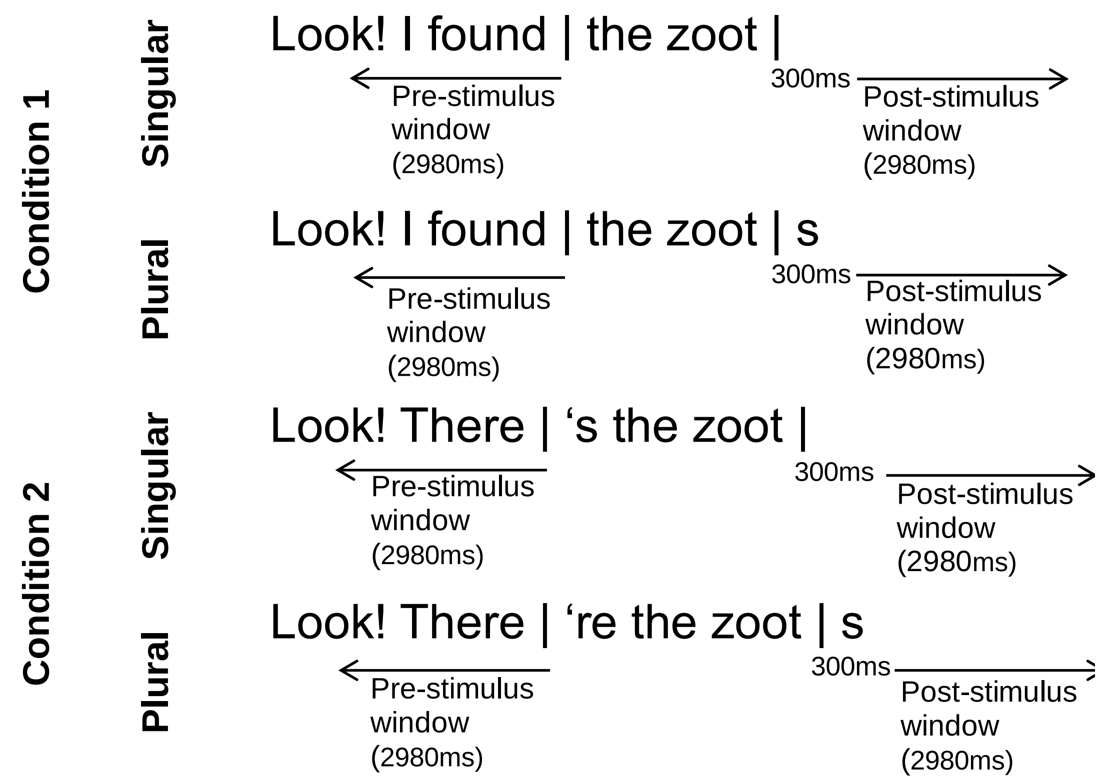
[9] If a participant looked at one or other image a proportion can be computed To calculate a difference score the therefore participant had to look at either the group or the individual at least once in both the pre-stimulus and post-stimulus windows. Trials for which this was not the case were excluded from the analysis. This leads to fewer data points which leads to a loss of power. It is, however, necessary for the analysis; a proportion cannot be calculated if the denominator is zero (which would be the case if the participant did not look at either the group or the individual) and a difference score cannot be calculated without a proportion from both windows. As such, difference Score could not be calculated for 41 out of 402 trials. Following this trial loss 1 participant failed to contribute data to all four conditions. Their data was excluded from this analysis. Consequently, this analysis was run with 354 trials (88%) across 32 participants (97%).

[10] The length of the analysis window is limited by the time between the end of trial and the offset of the nominal stem. Due to equipment failure some of the final frames of each trial were not recorded. The maximum number of milliseconds recorded for all trials was 11800. All trials were therefore cut at this point. As the labels are different lengths the offset of the nominal stem varies across items. The latest offset is 8520ms and so the gap between the latest offset and the end of the trial is 11800ms – 8520ms = 3280ms. Including a 300ms processing window, the maximum possible analysis window is therefore 3280ms – 300ms = 2980ms.

[11] This analysis was also run using the 1000ms window used by Paquette-Smith and Johnson (2016) and comparable results were obtained.

to 8320 for plural trials. During these periods participants had not been exposed to any number marked elements and the pre-stimulus windows were comparable across condition and number in that both included neutral sentential context. In both conditions, the post-stimulus window began 300ms after onset of the final number marked element (i.e. the offset of nominal stem). The 300ms account for the time it typically takes for a 3-year-old to initiate fixation on a visual target in response to an auditory stimulus (Fernald et al., 2008). The start of the post-stimulus window varies across items. The mean start point on singular trials is 8651ms and the mean start point on plural trials is 8574ms. During the post-stimulus window participants had been exposed to all number marked elements and again the windows were comparable across condition and number. The location of all windows of interest are represented diagrammatically in Figure 8 and indicated by grey shading in Figures 4-7 above (for the windows whose location varies across items the average is shown).

Figure 8: Analysis windows, Experiment 1



Results

The descriptive statistics are given in Table 2. All values used to calculate the means are within 2.5 standard deviations of the grand mean.

Table 2: Difference score means by number, condition, age and allomorph, proportional looks analysis, Experiment 1

	Mean (SD)	**Mean (SD) by condition**		**Mean (SD) by age**		**Mean (SD) by allomorph**	
		Cond1	Cond2	2yo	3yo	/s/	/z/
Sing.	-0.02 (0.18)	0.04 (0.28)	-0.05 (0.35)	-0.07 (0.22)	0.02 (0.14)	NA	NA
Plu.	0.04 (0.22)	0.05 (0.35)	0.05 (0.29)	0.02 (0.24)	0.06 (0.21)	0.01 (0.26)	0.06 (0.34)

To assess the impact of the independent variables on the difference score I used the lme4 package (Bates, Maechler, & Bolker, 2012) for R (R Core Team, 2019) to perform a linear mixed effects analysis. The model was created by entering number, condition and age (months) as fixed effects. The categorical variables condition and number were made numeric by assigning the value -0.5 to trials in Condition 1 and 0.5 to trials in Condition 2 and -0.5 to singular trials and 0.5 to plural trials. Age was centred by taking each participant's age in months and subtracting the mean. The variables were added to the model with interaction terms. As random effects the model had an intercept for participant and by-participant random slopes for the effect of number and condition.[12] [13] As each item occurs only twice in each condition (once singular and once plural) item was not included as a random effect. The output of this model is given in Table 3. As shown, there are no significant main effects or interactions. Finding no significant main effect of number suggests that participants did not look more at the group of objects after hearing plural sentences or more at the individual object after hearing singular sentences. The lack of significant interactions suggests this was the case in both conditions and across the age range tested.[14]

[12] Formula: model.full = lmer(diffscore ~ number * condition * age + (1 + number + condition | participant) , data=exp2.prop.looks)

[13] A by-participant random slope for the effect of age was not included as each participant has only one age.

[14] Due to the small amount of data collected it was not possible to run a comparable analysis with control items. As both familiar labels (chairs and spoons) took the allomorph it was also not possible to run a by-allomorph analysis for control items. This was also the case for all the analyses that follow.

Table 3: Results of a linear mixed effects model examining the effect of number, condition and age on difference score, proportional looks analysis, Experiment 1

Factor	Estimate	Std. Error	df	*t*	*p*
Intercept	0.0092	0.0243	28.8	0.376	0.7093
Number	0.0537	0.0535	32.0	1.003	0.3233
Condition	-0.0044	0.0494	73.2	-0.089	0.9291
Age	0.0028	0.0034	26.8	0.827	0.4153
Number:Condition	0.1595	0.0955	316.9	1.670	0.0959
Number:Age	-0.0008	0.0075	30.2	-0.112	0.9119
Condition:Age	0.0105	0.0069	67.9	1.520	0.1331
Number:Condition:Age	0.0129	0.0133	309.9	0.965	0.3355

To test whether participants were more sensitive to the /s/ allomorph than the /z/ allomorph, I created another model which was run only with plural trials.[15] As participants of different ages may respond differently to the two allomorphs, this model took both allomorph and age as fixed effects and allowed these variables to interact. Allomorph was made numeric by the assigning the value -0.5 to trials on which the novel label was inflected /z/and 0.5 to trials on which the label was inflected /s/. Age was centred by taking each participant's age in months and subtracting the mean. As random effects the model had an intercept for participant as well as a by-participant random slope for the effect of allomorph. Again, as shown in Table 4, the results show no significant main effects or interactions. This suggests that, regardless of age, participants spent no more time looking at the group after hearing plural labels inflected with the /s/ allomorph than plural labels inflected with the /z/ allomorph.

Table 4: Results of a linear mixed effects model examining the effect of allomorph and age on difference score, proportional looks analysis, Experiment 1

Factor	Estimate	Std. Error	df	*t*	*P*
Intercept	0.03716	0.03767	27.33	0.986	0.333
Age	0.003305	0.005324	26.46	0.621	0.540
Allomorph	-0.020392	0.071160	148.48	-0.287	0.775
Age:Allomorph	-0.011048	0.010053	149.06	-1.099	0.274

In the analysis above, 'attention' was conceptualised as the amount of time spent looking at the target as a proportion of time spent looking at the target and competitor combined. It is possible, however, for attention to be manifested in other ways. For

[15] Formula: lmer(diffscore ~ age * allomorph + (1 + allomorph | participant), data=plu)

example, Schafer and Plunkett (1998) measured attention in terms of continuous looking and calculated whether participants' longest look to the target occurs post-stimulus or pre-stimulus. The following section presents the results of an analysis using this measure as the dependent variable.

3.4.3 Longest looks

Analysis

This analysis is based on the rationale that if the amount of attention a participant pays to the target increases post-stimulus there will be a longer period of sustained attention to the target following the stimulus than preceding the stimulus (Schafer & Plunkett, 1998). That is, the longest look-to-target post-stimulus should be longer than the longest look-to-target pre-stimulus.

The dependent variable in this analysis is calculated by subtracting the longest look-to-individual from the longest-look-to-group in both the pre-stimulus and post-stimulus windows. This creates two difference scores. The pre-stimulus difference score is then subtracted from the post-stimulus difference score to yield a third difference score. The higher the difference score the greater the post-stimulus increase in the length of the longest look-to-group relative to the increase in the length of the longest look-to-individual. The lower the difference score the greater the post-stimulus increase in the length of the longest look-to-individual relative to the increase in the length of the longest look-to-group. If participants respond appropriately to the plural stimuli, the longest look-to-group post-stimulus will be longer than the longest look-to-group pre-stimulus. If they respond appropriately to the singular stimuli, the longest look-to-individual post-stimulus will be longer than the longest look-to-individual pre-stimulus. If these patterns hold, difference scores will therefore be significantly higher than zero on plural trials and significantly lower than zero on singular trials. The windows were of the same length and in the same position as those used in the proportional looks analysis described above (see section 3.4.2 for details).

Results

The descriptive statistics are shown in Table 5. Any values that were greater than or equal to 2.5 standard deviations above or below the mean were excluded from further analysis.[16]

Table 5: Difference score (ms) means by condition, age and allomorph, longest looks analysis, Experiment 1

	Mean (SD)	Mean (SD) by condition		Mean (SD) by age		Mean (SD) by allomorph	
		Cond1	Cond2	2yo	3yo	/s/	/z/
Sing.	-2.06 (55.6)	19.65 (75.6)	-13.60 (80.3)	-9.59 (59.8)	3.80 (53.1)	NA	NA
Plu.	2.63 (66.1)	-6.25 (82.4)	8.33 (75.1)	6.65 (65.3)	-0.49 (68.4)	-9.03 (106.1)	18.65 (71.8)

To assess the impact of the independent variables on the difference score I used the lme4 package (Bates et al., 2012) for R (R Core Team, 2019) to perform a linear mixed effects analysis. The model was created by entering number, condition and age (months) as fixed effects. The categorical variables condition and number were made as numeric as before and age was centred. The variables were added with interaction terms. As random effects the model had an intercept for participant and a by-participant random slope for the effect of number and condition.[17] The results of this model is shown in Table 6.

Table 6: Results of a linear mixed effects model examining the effect of number, condition and age on difference score, longest looks analysis, Experiment 1

Factor	**Estimate**	**Std. Error**	**df**	***t***	***P***	
Intercept	-0.4493	7.5169	85.6	-0.060	0.9525	
Number	2.2039	16.0641	42.0	0.137	0.8915	
Condition	-7.7041	15.0124	133.4	-0.513	0.6087	
Age	-0.5740	1.0548	81.1	-0.544	0.5878	
Number:Condition	42.4714	29.3685	359.6	1.446	0.1490	
Number:Age	-0.2389	2.2585	40.2	-0.106	0.9163	
Condition:Age	4.0981	2.1077	127.6	1.944	0.0541	.
Number:Condition:Age	2.1776	4.1165	355.8	0.529	0.5971	

[16]Ten out of 402 trials were excluded as a result. Following this trial loss, 1 participant failed to contribute data to all four conditions. Their data was excluded (9 trials). A further 9 trials were excluded from the remaining participants and this analysis was therefore run with 384 trials (96%) across 32 participants (97%). A follow-up analysis with these data included yielded comparable results. The means reported are those with these data excluded.
[17] Formula: model.full = lmer(diffscore ~ number * condition * age + (1 + number + condition | participant), data=exp2.long.looks)

The results do not show a significant main effect of number or any significant interactions involving number. This suggests that participants did not look longer at the group after hearing plural stimuli or at the individual after hearing singular stimuli. In this way, this analysis provides no evidence that participants are sensitive to number marking in this context. The results do, however, show a near-significant interaction between condition and age (p = .05). To explore this further I ran t-tests comparing 2- and 3-year-olds' difference scores to zero in both Condition 1 and Condition 2. The results were not significant. I also compared 2- and 3-year-olds' difference scores to each other in Condition 1 and Condition 2. Again the results were not significant. To explore the relationship between age and condition further the sample was split into 4 groups based on age in months. This was done to assess the possibility that the developmental differences found involved only the youngest 2-year-olds or oldest 3-year-olds (rather than 2- and 3-year-olds in general). Age group 1 contained the eight youngest participants (24-30 months), age group 2 the next 8 youngest (31-36 months), age group 3 the next 8 youngest (38-34 months) and age group 4 the 8 oldest participants (43-46 months). The mean difference scores by age group and condition are given in Table 11. Planned comparisons showed no significant differences between age groups in Condition 2. In Condition 1, age group 1's difference score (48.00ms) was significantly higher than age group 2's (-23.08ms). A positive difference score indicates an increase in attention to the group and this finding therefore suggests that in Condition 1, regardless of whether they heard singular and plural stimuli, the youngest participants tended to look at the group. It is not clear why this was not the case in Condition 2.

Table 7: Difference score (ms) means by condition and age group, longest looks analysis, Experiment 1

	Age group 1	**Age group 2**	**Age group 3**	**Age group 4**
Condition 1 mean (sd)	48.00 (137.9)	-23.08 (158.7)	7.44 (129.8)	-11.06 (144.6)
Condition 2 mean (sd)	-25.10 (147.1)	-8.78 (138.13)	22.40 (139.5)	-8.73 (150.6)

To test whether participants' sensitivity to the plural stimuli varied according to allomorph, I created another model to be run with plural trials only.[18] As before, this model took allomorph and age (added with an interaction term) as fixed effects. Allomorph was made numeric and age was centred as before. As random effects the model had an intercept for participant as well as a by-participant random slope for the effect of allomorph. As shown in Table 8, the results indicate no significant main effects

[18] Formula: lmer(diffscore ~ age * allomorph + (1 + allomorph | participant), data=plu)

or interactions. This suggests that in both conditions and regardless of age, participants looked no longer at the group after hearing plural labels inflected /s/ than after hearing plural labels inflected /z/. In this way the results of the longest looks analysis support those of the proportional looks analysis in indicating that participants were not sensitive to the stimuli.

Table 8: Results of a linear mixed effects model examining the effect of allomorph and age on difference score, longest looks analysis, Experiment 1

Factor	Estimate	Std. Error	df	*t*	*P*
Intercept	1.1855	12.0433	38.12	0.098	0.922
Age	-0.6358	1.6906	36.56	-0.376	0.709
Allomorph	-26.9461	23.2201	45.89	-1.160	0.252
Age:Allomorph	-3.9194	3.2575	44.15	-1.203	0.235

3.4.4 Additional analyses

Other possible dependent variables were reaction time and shift probability. Due to data loss, however, it was not possible to run these analyses with this data. Details are given in Appendix 4.

3.5 Discussion

The results of this experiment show that in this context 2- and 3-year-olds do not follow grammatical number. This is the case both when number is marked on the noun and the verb (e.g. 'There's the *zoot*!' or 'There're the *zoots*!') and when it was marked on the noun only (e.g. 'I found the *zoot*!' or 'I found the *zoots*!'). Apart from the near-significant interaction between condition and age discussed, no significant developmental differences were found. In this way, this experiment provides no evidence that 2- and 3-year-olds are sensitive to grammatical number. This is in contrast to other reported findings showing young children succeed in similar tasks from the age of 24 months when number is marked on the determiner, noun and verb and from the age of 30 months when number is marked on the noun only (see Section 1.3 for details). Why, then, did the participants fail to show this same sensitivity here? One explanation concerns the nature of the potential referents. In particular, as discussed in Section 3.1, in previous studies the potential referents belonged to two different categories. For example, participants might have been shown a group of honey spoons and a single lemon juicer. In the present study, however, participants were shown potential referents belonging to the same category, e.g. a group of knife guards and a single knife guard. As discussed, the present task more accurately reflects the task children face in the real world. In this way, previous studies may overestimate children's ability to follow grammatical number.

An alternative explanation is that the other aspects of the methodology used in the present study contributed to the participants' behaviour. For example, the speaker was facing away from the participants when she produced the sentences. It is possible that participants did not pay attention to what she said because it was not clear that her speech was meant 'for them'. However, the actor began each trial by facing the child and used child directed speech throughout. These factors should have facilitated the participant's understanding that the actor was attempting to engage with them. Furthermore, in previous studies (for example Jolly & Plunkett, 2008; Kouider et al., 2006; Lukyanenko, 2011) participants successfully attended to the sentences even though no speaker was visible. It is not clear why children would attend to a disembodied voice, but not to the voice of a speaker facing away from them. A second factor concerns the number of times the number marked labels were repeated. In particular, in both Kouider et al. (2006) and Jolly and Plunkett (2008)'s experiments participants heard the sentence containing the novel label and also a repetition of the label in isolation. Similarly, in Paquette-Smith and Johnson (2016)'s recent experiment participants heard the novel label presented in two number marked sentences, e.g. 'There are nice *blickets.* Can you find the *blickets*?'. In the present study, however, participants heard the number marked novel label only once. It is possible that in the previous studies hearing the novel label repeated highlighted the number marking and encouraged participants to follow this cue.

Another potential explanation for why participants did not attend to number marking in this context concerns the nature of the potential referents. In particular, recall that in this experiment the members of the group and the individual both belonged to the same category, e.g. a group of knife guards and a single knife guard. As such, the task participants were faced with was determining which one of two displays a speaker was using a novel label to refer to, not what category of object the label referred to per se. I have characterised this as being speaker reference rather than word reference. In this way, although participants were still solving a referential problem, they did not have to disambiguate the novel label.

The experiment used referents of two different colours to try to give a plausible reason for why the speaker was referring to only one of the objects (e.g. it was the purple one/s in particular she was interested in). If this failed, and participants did not read the task as involving the disambiguation of even speaker reference, there should be no difference between the current results and the results of a condition where objects of the same colour were used. This represents an avenue for future research. A further issue with this design is that on plural trials labels (i.e. when they heard ‘I found the *zoots*!’ in Condition 1 or when they heard ‘There’re the *zoots*!’ in Condition 2) participants may have interpreted the label as referring to the combination of the individual and group of objects. That is, the individual and the group could have been seen as one large group of *zoots*. If so, a reasonable response would have been to look between the two displays rather than showing a preference for the group. In this way, it is possible that this feature of the design masked sensitivity to the number cue, at least on plural trials. One factor that speaks against this interpretation, though, is the looking behaviour of the participants on plural trials. In particular, as Figure 5 and Figure 7 show, in the post-stimulus window of plural trials participants tended to look at the speaker, rather than indiscriminately between the displays. This suggests that they were not interpreting the plural labels as referring to the two displays collectively. This was also the case on singular trials (see Figure 4 and Figure 6). In this way, participants behaved comparably on singular and plural trials. The clearest conclusion is therefore that participants were not sensitive to number marking in this task.

The following chapter presents an experiment that explores how 2- and 3-year-old children use gesture to solve the reference problem. In particular, it investigates whether they are more attentive to the combination of body orientation and pointing than to body orientation alone.

Chapter 4 Experiment 2: Gesture

4.1 Aims and Predictions

As discussed in Section 1.2, by the time they are 12 months old, infants are able to follow pointing gestures (Stephens & Matthews, 2014). By the time they are 18 months old, they are able to use pointing to inform their referential choices and, over the course of naming episodes, assign novel labels to the consistent referents of pointing gestures (Briganti & Cohen, 2011). At 12 (and also 18) months old, young children are able to follow body orientation cues and will follow the orientation of an actor's head and body to a target (Deák et al., 2000; Tomasello et al., 2007). There is, however, no evidence that this is the case in tasks that have a referential component. One of the aims of the present experiment is to determine whether older learners (i.e. 2- and 3-year-olds) can follow body orientation in a referential task. As discussed in Section 1.2, this has implications for children's ability to learn from overseen interactions.

Furthermore, as discussed in Section 1.2, previous studies suggest that children of these ages are more attentive to stimuli indicated by gestures comprising two elements than to those indicated by stimuli comprising a single element. However, the most relevant previous study (Grassmann & Tomasello, 2010) conflates the amount of information a cue gives with the quality of information that cue gives. Specifically, it compared a non-communicative pointing gesture with a communicative looking and pointing gesture. The present experiment compares two-element gestures and one-element gestures when the communicative context the gestures are presented in is controlled for. This issue is considered developmentally by testing children of a wide age range, namely 2- and 3-year-olds. The reason for this is that developmental differences may provide an explanation for any observed preference for the two-element gesture (see Section 1.2.2 above). My second research question is therefore as follows:

RQ2: Do 2- and 3-year-olds follow body orientation? If so, do they follow the combination of body orientation and pointing more than body orientation alone? Does this change across development?

This question is addressed by showing participants two displays (e.g. a yellow knife guard and a purple knife guard) and while a speaker says a sentence that could refer to either (e.g. 'I found a *zoot*!'). The speaker then either faces (Condition 1) or faces and points to (Condition 2) one of the displays. As she produces the gesture her back is to the participant. If participants are sensitive to the gestures they should pay more attention to the display she gestures at than to the other display. If they are more sensitive to body orientation and pointing than to body orientation alone they should do so more in Condition 2 than Condition 1. The main variable of interest is therefore condition and it is predicted that 2- and 3-year-olds will: (a) follow both types of gesture; and (b) follow body orientation and pointing more than body orientation alone. As described above, a second variable of interest is age.

Like in Experiment 1, Experiment 2 used potential referents belonging to the same category. For example, when hearing 'I found the *zoot*!' participants were presented with two knife-guards (of different colours). As discussed previously, the task is referential, but concerns speaker reference rather than word reference. This decision was made because in Experiment 3 I intend to look at children's use of the body orientation cue in and out of alignment with the number cue used in Experiment 1. As such, it is important to ensure learners could use the body orientation cue in isolation to discern speaker reference. In this way, as well as addressing the research questions discussed previously, Experiment 2 also acts as a pre-test for Experiment 3.

4.2 Methodology

4.2.1 Participants

Forty-one 2- (n = 21, 10 girls) and 3-year-olds (n = 20, 11 girls) participated in this experiment (*M* = 3;0, Range = 1;11-3;10). None of these participants had taken part in Experiment 1. Participants came from monolingual English-speaking homes (n = 39) or homes where English was the dominant language (n = 2). Participants were primarily recruited through local online parenting forums and university mailing lists. Parents gave informed consent on behalf of their children in all cases (see Appendices 1 and 2). The experiment was approved by the University of Edinburgh Philosophy, Psychology and Language Sciences Ethics Committee. As compensation for their time, parents were given a £2.50 voucher for a local coffee shop and children were given a sticker book. Twelve additional children (four 2-year-olds and eight 3-year-olds) participated in the experiment, but were excluded from analysis for the following reasons: restlessness (n = 6), parental interference (n = 1), equipment failure (n = 1), experimenter error (n = 3) and parentally-reported language delay (n = 1).

4.2.2 Design

The main independent variable was condition: body orientation (Condition 1) versus body orientation and pointing (Condition 2). This was varied within subjects.

As in Experiment 1, 4 versions of the experiment were created to counter-balance two factors for each label: (a) target position and (b) the object label pairing.[19] The colour of the target and the competitor were balanced as in Experiment 1. Direction of turn was yoked to labels such that half of the labels always occurred with a clockwise turn and half with an anti-clockwise turn. These lists are reproduced in Appendix 3. Participants were assigned to lists pseudo-randomly with the constraint that equal numbers of participants completed each one.[20]

As before, participants completed two different types of trial: experimental trials in which unfamiliar objects were presented with novel labels (e.g. 'There are the *zoots!*') and control trials in which familiar objects were presented with their conventional labels (e.g. 'There are the spoons!'). All participants completed 16 experimental trials and 4 control trials.

4.2.3 Materials

Audio stimuli

Twelve sentences served as audio stimuli. These were identical to those used in Condition 1 (nominal number marking only) of Experiment 1.

Visual stimuli

The images used were the same as those in Experiment 1. Rather than presenting a single image alongside a group of images, however, single images were presented with another single image and groups of images were presented with another group of images (see Figure 9).

4.2.4 Trial procedure

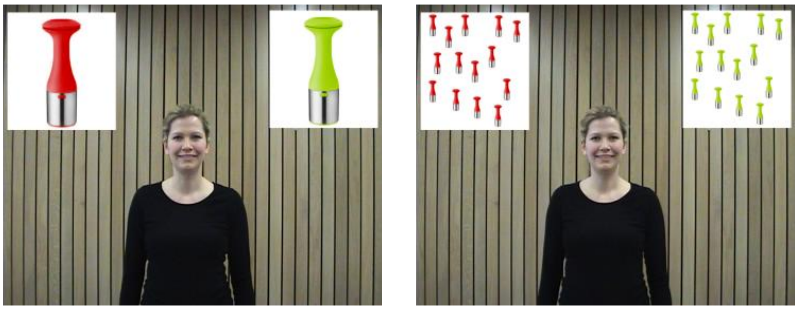
In each trial two images were displayed. These images were either two individual objects or two groups of objects presented against a white rectangular background. The objects were always the same category and different colours. On all trials, the grammatical number of the sentence and the number of objects shown on each display were matched (i.e. singular sentences always co-occurred with two pictures of a single object and plural sentences always co-occurred with two pictures of a group of objects). The rectangles were positioned as before. As in Experiment 1, these images were super-imposed above a female actor standing in front of a wooden slatted wall such that

[19] The object-label pairing was varied for unfamiliar objects and labels only; familiar objects were always presented with their conventional labels.

[20] Following participant loss, the final proportions were as follows: 12/41 participants completed List A; 9/41 completed List B; 10/41 completed List C; and 10/41 completed List D.

the images appeared to be on the wall with the actor standing between them (see Figure 9).

Figure 9: Visual stimuli in situ, Experiment 2. Left panel: individual objects, right panel: groups of objects



The trials progressed much as in Experiment 1, except that rather than looking straight ahead after turning to face the wall, the actor either turned to face (Condition 1) or turned to face and pointed at (Condition 2) one of the displays. Details of within-trial timings are found in Figure 10.

4.2.5 Session procedure

As Experiment 1.

4.3 Coding

The same coding procedure was used as in in Experiment 1.[21] As before, a second coder independently coded complete data sets from 8 participants (20% of the useable data) Of these, 4 were randomly selected from the 2-year-old participants and 4 from the 3-year-old participants. Agreement between the coders was high. When all four coding categories were used the mean agreement score was of 82% (*SD* = 4.4, Cohen's κ = 0.75). Traditionally, this is interpreted as indicating a substantial degree of agreement (Landis & Koch, 1977). When the 'speaker' and 'elsewhere' codes were collapsed, mean agreement between coders increased to 88% (*SD* = 4.0, Cohen's κ = 0.81). This indicates a 'near perfect' degree of agreement (Landis & Koch, 1977) When the coders disagreed, the first coder's decision was used. The data were then filtered such that any

[21] As can be seen in Figure 10, when the speaker pointed to the displays her hand was quite close to the display. In theory, it may therefore have been difficult to reliably determine whether the participant was fixated on the speaker's hand or on the display she pointed to. In reality, it was possible to distinguish between the two. Indeed, one participant had fixation points corresponding to both and the display fixation point was notably different from the hand fixation point. In future studies, though, this potential issue could be avoided by increasing the distance between the hand and the displays.

trial on which greater than or equal to 20% of frames were coded ‘elsewhere’ was excluded from analysis.[22]

Figure 10: Trial progression, Experiment 2

Time (ms)	Audio	Visual			
		Actor	**Images**	**Condition 1: body orientation**	**Condition 2: body orientation + pointing**
0		Facing camera	Absent		
1000	Boing!	Facing camera	Present		
2080	Gasp!	Facing camera	Present		
3160	‘Let’s see!’	Turning to wall	Present		
4000		Turning to wall	Present		
5560	‘Look!’	Facing wall	Present		
6000		Facing wall	Present		
7280	‘I found...	Facing wall	Present		
8000	... the *zoot*!’	Starts gesturing	Present		
9000		Gesturing	Present		
10000		Gesturing	Present		
11000		Gesturing	Present		
12000		Stops gesturing	Present		
13000	Boing!	Turning to camera	Absent		
14000		Turning to camera	Absent		
15000		Turning to camera	Absent		
16000		Facing camera	Absent		

[22] Comparable results were found when these trials were included. The total number of trials excluded was 279 out of 820. This equates to approximately 34% of the useable data collected. Following this loss, all participants contributed at least one trial per condition. One additional trial was removed due to equipment failure leaving 540 trials in total. This comprises 432 experimental trials and 108 control trials. The following analyses are conducted with the experimental trials only.

4.4 Analysis and Results

4.4.1 Eye movements

The analysis begins by considering participants' gaze-direction data to four visual areas: target, competitor, speaker and elsewhere. Overall, the eye-movement data suggest participants were engaged in the task and paying attention to events on the screen: They looked at the stimuli when they first appeared and at the speaker when she spoke and moved. Furthermore, they provide initial evidence of a difference between the efficacy of the gestures and suggest that while the body orientation and pointing gesture was effective in directing participants' attention, the body orientation only gesture was not.

To illustrate, as shown in Figure 11 (Condition 1) and Figure 12 (Condition 2) there is a peak in looks to both displays (i.e. to both target and competitor) when they first appear on screen (1000ms). As one would expect, there is no preference for either target or competitor at this stage. This peak is then followed by a sustained period of looks-to-speaker as she 'hears' the objects appear and turns to look at them (between approximately 2250ms and 4750ms). Looks to the stimuli and speaker then level out until after the actor begins gesturing at 8000ms. At this point, looking behaviour begins to differ across conditions. Specifically, in Condition 2 (Figure 12), from around 8500ms there is a clear increase in looks-to-target and corresponding decrease in looks-to-competitor and looks-to-speaker. This is what one would expect to see if the gesture is effective in directing attention on to the target and away from the competitor. In Condition 1 (Figure 11) however, both looks-to-target and looks-to-competitor decreased at this point, while looks-to-speaker increased. This pattern suggests that unlike in Condition 2, the gesture in Condition 1 was not successful in directing participants' attention onto the target. The following sections explore whether these initial observations are supported by statistical analyses.

Figure 11: Looks to target (red), competitor (green), speaker (blue) and elsewhere (purple), Experiment 2, Condition 1 (body orientation)

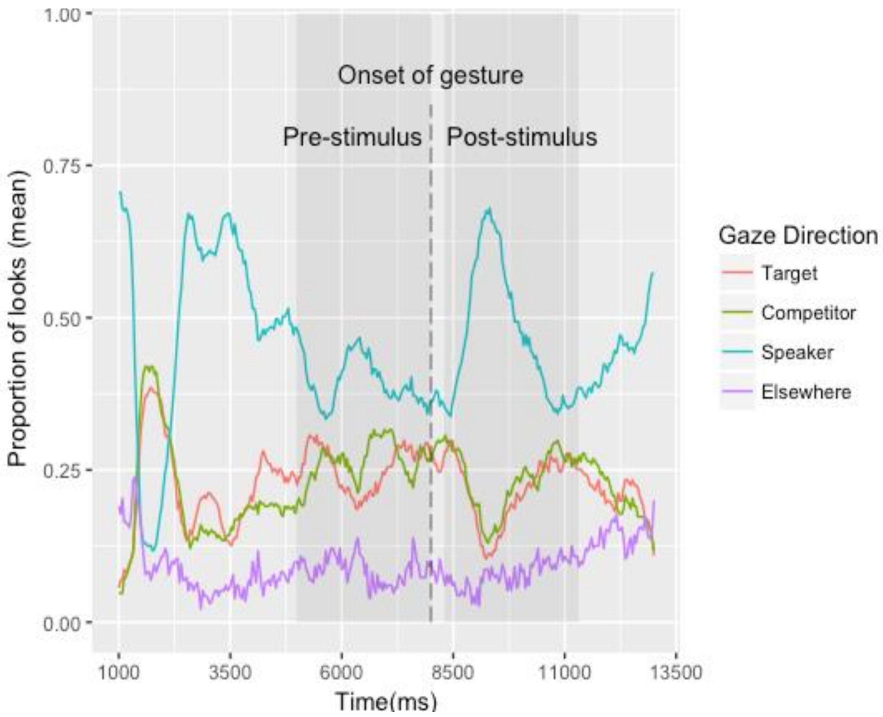
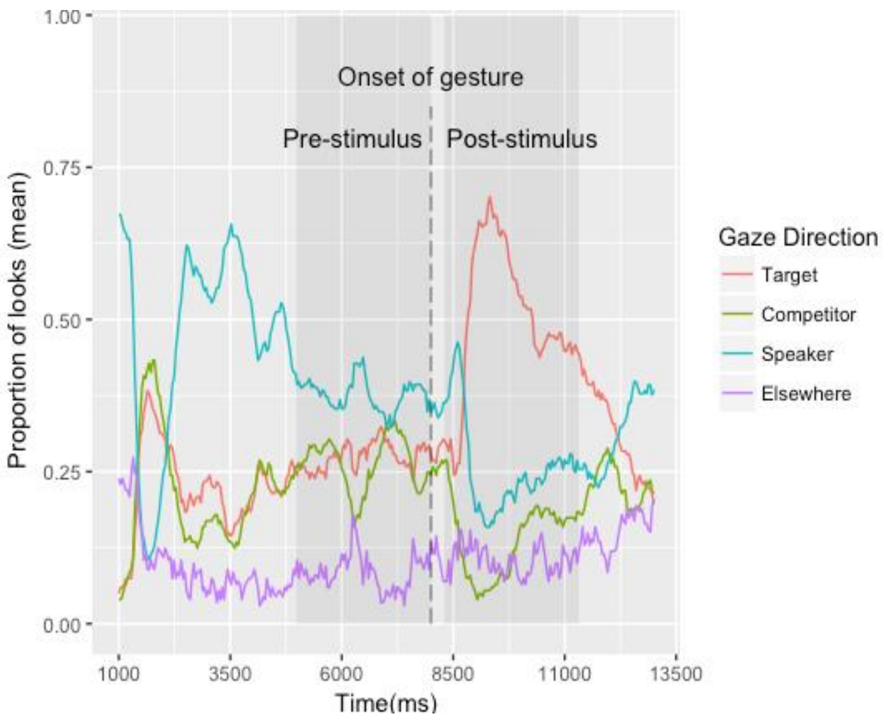


Figure 12: Looks to target (red), competitor (green), speaker (blue) and elsewhere (purple), Experiment 2, Condition 2 (body orientation and pointing)



4.4.2 Proportional looks

Analysis

As discussed in Section 4.1, the central prediction is that participants will pay more attention to the target following the stimuli in Condition 2 than the stimuli in Condition 1. The main independent variable is therefore condition. As I am interested in developmental differences, a second variable of interest is age.

Similarly to Experiment 1, the first dependent measure is based on proportion of looks-to-target (Hollich et al., 2000; Kouider et al., 2006; Paquette-Smith & Johnson, 2016). Proportion of looks-to-target refers to the amount of time participants spend looking at the target (i.e. the display the actor gestured toward) as a proportion of looks-to-target and looks-to-competitor combined. Following previous preferential looking studies exploring young children's responses to social eye-gaze (Hollich et al., 2000; Paquette-Smith & Johnson, 2016), this is measured before and after the onset of the gesture (7000ms). In keeping with the recommendation for this kind of methodology (Fernald et al., 2008), the post-stimulus window begins 300ms after the onset of the stimulus i.e. at 8300ms. The duration of the analysis window is set at 3000ms for both the pre- and post-stimulus windows. This duration was chosen for two reasons: (a) it is between the windows used in the two studies with the most similar methodologies (i.e. the 1000ms window used by Paquette-Smith and Johnson (2016) and the 6000ms windows used by Hollich et al. (2000)) and (b) it is very close to the duration of the analysis windows for this dependent variable in Experiment 1 (i.e. 2980ms, see Section 3.4.2 for details).[23] The pre-stimulus window therefore ran from 5000ms-8000ms and the post-stimulus window from 8300-11300ms. These windows are indicated by the grey shading in Figure 11 and Figure 12 above. Within each analysis window, the amount of time participants spent looking at the target was calculated as a proportion of the time spent looking at the target and the competitor combined.[24] The post-stimulus proportion was then subtracted from the pre-stimulus proportion to generate a difference score. A difference score that is significantly greater than zero indicates that the proportion of looks-to-target is higher post-stimulus than pre-stimulus. A difference score that is significantly lower than zero indicates that the proportion of looks-to-target is lower post-

[23] This analysis was also run with the longest possible window (4700ms) and a much shorter window (1000ms). The results of each analysis were comparable.

[24]To calculate a difference score, participants therefore had to have looked at either the target or the competitor at least once in both the pre- and post-stimulus windows. This was not the case on 36 out of 432 trials. These trials were excluded from this analysis. Following this loss, 1 participant failed to contribute data to both conditions. Their data was excluded from this analysis. This analysis was therefore run with the data from the remaining 40 participants only (98%). This equates to 391 trials (91%).

stimulus than pre-stimulus. If the stimuli are effective in directing attention to the target, the difference scores should be higher than zero in both conditions. If body orientation and pointing directs attention more effectively than body orientation alone, the difference score should be higher in Condition 2 than in Condition 1.

Results

The descriptive statistics are given in Table 9. All values used to calculate the means are within 2.5 standard deviations of the grand mean.

Table 9: Difference score means by condition and age, proportional looks analysis, Experiment 2

	Mean (SD)	**Mean (SD) by age**	
		2yos	3yos
Cond.1	-0.04 (0.27)	0.06 (0.20)	-0.14 (0.33)
Cond. 2	0.25 (0.20)	0.26 (0.23)	0.24 (0.16)

To assess the impact of condition and age on difference score I used the lme4 package (Bates et al., 2012) for R (R Core Team, 2019) to perform a linear mixed effects analysis. As fixed effects the model had condition and age (months). Condition was made numeric by assigning the value -0.5 to trials in Condition 1 and 0.5 to trials in Condition 2. Age was centred as in Experiment 1. These variables were added with interaction terms. As random effects, the model included an intercept for participant and a by-participant random slope for the effect of condition.[25] Results of the model are given in Table 10. As shown, the model suggests that difference score is significantly affected by condition and not affected by age.[26]

Table 10: Results of a linear mixed effects model examining the effect of condition and age on difference score, proportional looks analysis, Experiment 2

Factor	**Estimate**	**Std. Error**	**df**	***t***	***P***	
Intercept	0.1280	0.0260	39.07	4.925	<.001	***
Condition	0.2617	0.0478	71.75	5.472	<.001	***
Age	-0.0029	0.0039	36.99	-0.760	0.452	
Condition:Age	0.0101	0.0071	68.70	1.430	0.157	

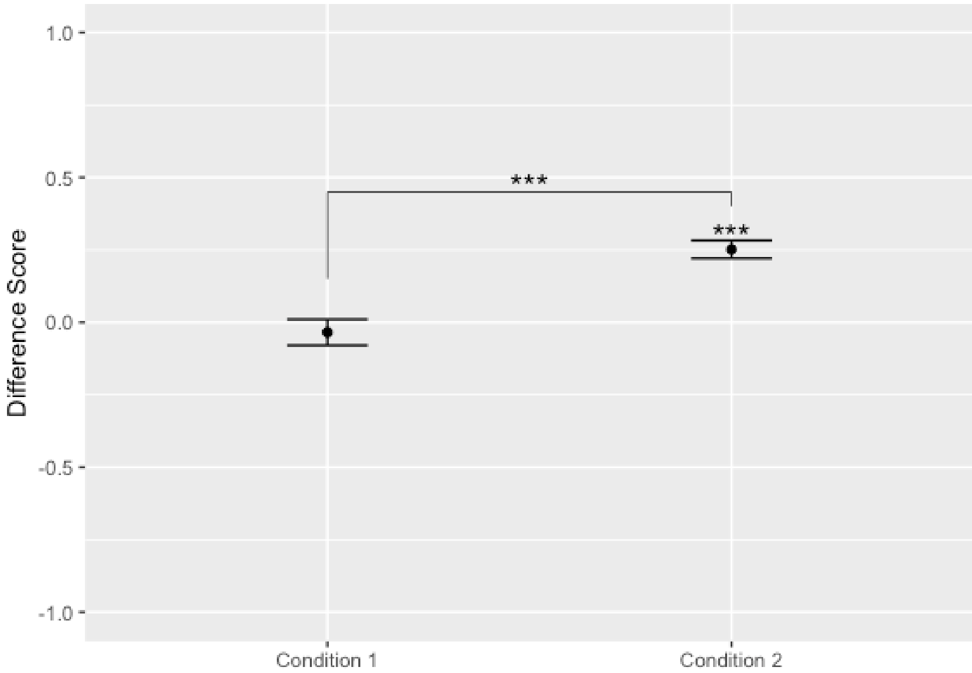
[25] Formula: model.full = lmer(diffscore ~ condition * age + (1 + condition | participant), data=exp1.prop.looks)

[26] Throughout this thesis I follow the convention that p-values of less than .001 are marked '***', those of less than .01 are marked '**', those of less than .05 are marked '*' and those of less than 1 are marked '.'.

To confirm that age does not significantly affect difference score, I created a second model that did not include age as a fixed effect.[27] The models were then compared using ANOVA. The results showed that the fit of the full model was no better than the fit of the reduced model (log likelihood: -243 versus -244, ($\chi2(2) = 2.38$, $p = .31$). I therefore continue with the full model. To confirm that condition does significantly affect difference score, I created a third model with condition and its associated slopes removed.[28] When the two models were compared, I found that the model including condition had a better fit than the reduced model (log likelihood: -244 versus -260, ($\chi2(2) = 31.56$, $p < .001$). This confirms the finding that difference score is significantly affected by condition. In particular, as illustrated in Figure 13, the difference score in Condition 2 ($M = .25$) is significantly higher than the difference score in Condition 1 ($M = -0.4$). In this way, the results show a post-stimulus increase in the proportion of looks-to-target in Condition 2, and no such increase in Condition 1. This suggests that while the body orientation and pointing gesture was successful in drawing participants' attention to the target, the body orientation only gesture was not. This analysis therefore provides evidence that children are sensitive to body orientation and pointing, but provides no evidence of participants' sensitivity to body orientation alone.

[27] Formula: lmer(diffscore ~ condition + (1 + condition | participant), data=exp1.prop.looks)
[28] Formula: lmer(diffscore ~ age + (1 | participant), data=exp1.prop.looks)

Figure 13: Difference score by condition, proportional looks analysis, Experiment 2. Error bars show standard error



As was the case in Experiment 1, I also ran an analysis based on longest looks-to-target. This was done with the aim of (a) confirming participants' sensitivity to the body orientation and pointing gesture and (b) revealing any sensitivity to the body orientation gesture. The results of the longest looks analysis are presented in the following section.

4.4.3 Longest looks

Analysis

This analysis considers longest looks-to-target and is based on the rationale that if participants are sensitive to the stimulus the longest look to target should occur post-stimulus rather than pre-stimulus (Schafer & Plunkett, 1998). To determine whether the gestures increase participants' attention specifically to the target (rather than to both the target and the competitor), the analysis considers both longest looks-to-target and longest-looks-to-competitor. Specifically, the dependent variable is calculated by subtracting the longest look-to-competitor from the longest look-to-target in both the pre-stimulus and post-stimulus windows. This yields two difference scores. The pre-stimulus difference score is then subtracted from the post-stimulus difference score to give a third difference score. The higher the difference score the greater the post-stimulus increase in the length of the longest look-to-target relative to the increase in the length of the longest look-to-competitor. The lower the difference score the greater the post-stimulus increase in the length of the longest look-to-competitor relative to the increase in the length of the longest look-to-target. If the gestures are successful in directing attention, the difference score should be significantly greater than zero in both conditions. If body orientation and pointing orients attention more effectively than body orientation alone, the difference score in Condition 2 should be significantly higher than difference score in Condition 1. As in the previous analysis, the pre-stimulus window ran from 4000ms-7000ms and the post-stimulus window from 7300-10300ms.[29]

Results

The descriptive statistics are given in Table 11. Any values that were greater than or equal to 2.5 standard deviations above or below the mean were excluded from analysis.[30]

Table 11: Difference score means (ms) by condition and age, longest looks analysis, Experiment 2

	Mean (SD)	**Mean (SD) by age**	
		2yos	3yos
Cond.1	-12.6 (80.4)	2.6 (78.4)	-28.4 (82.5)
Cond. 2	81.8 (68.5)	90.5 (72.8)	72.6 (64.2)

I assessed the impact of the independent variables on difference score by using the lme4 package (Bates et al., 2012) for R (R Core Team, 2019) to perform a linear mixed effects analysis. The model was constructed in the same way as before: the fixed effects were condition and age (months), and the random effect was participant. The model also included a by-participant random slope for the effect of condition. The fixed effects were allowed to interact.[31] The results of this model are given in Table 12.

[29]I also ran the analysis with the longest possible window (4700ms) and a much shorter window (1000ms). The results were comparable.

[30] Ten out of 432 trials were excluded in this way and the following analysis was run with the remaining 422 trials only (98%). Following this trial loss, all participants still contributed data to both conditions. Follow up analyses in which these data were included yielded comparable results.

[31] Formula: model.full = lmer(diffscore ~ condition * age + (1 + condition | participant), data=exp1.long.looks)

Table 12: Results of a linear mixed effects model examining the effect of condition and age on difference score, longest looks analysis, Experiment 2

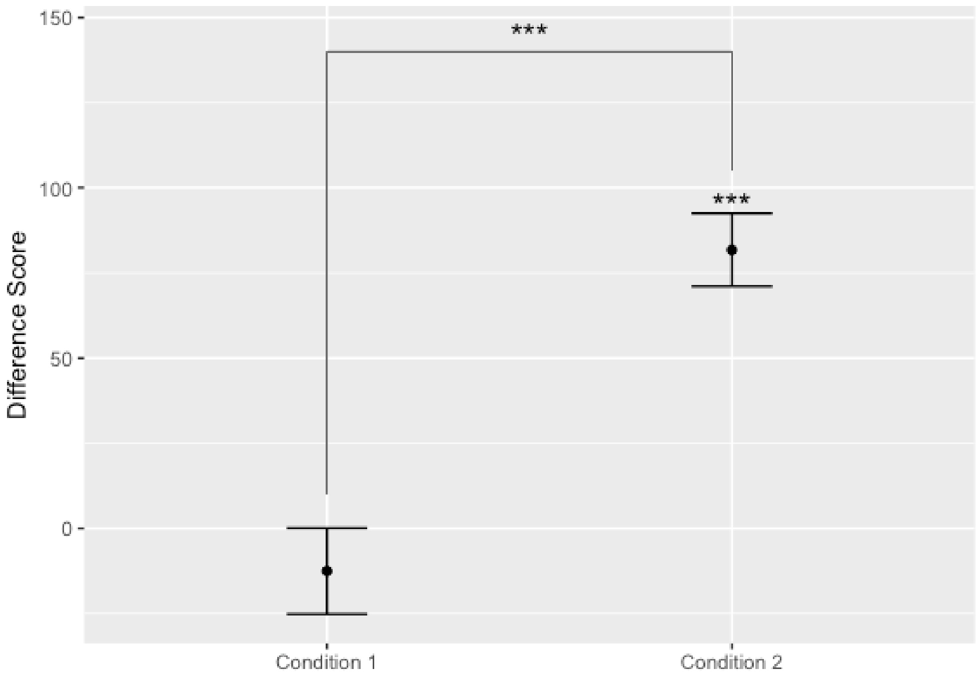
Factor	Estimate	Std. Error	df	*t*	*P*	
Intercept	40.7616	8.1122	47.03	5.025	<.001	***
Condition	87.7129	16.0037	47.28	5.481	<.001	***
Age	-1.4833	1.2033	45.08	-1.233	0.224	
Condition:Age	0.1389	2.3731	45.38	0.059	0.954	

As shown, this model indicates that condition, but not age, has a significant impact on difference score. To confirm this is the case, I created a reduced version of the model that did not included age as a fixed effect.[32] The two models were then compared using ANOVA. The results showed that the full model was no better a fit than the reduced model (log likelihood: -2711 versus. -2712, ($\chi 2(2) = 1.55$, $p = .46$). I therefore proceeded with the full model. To confirm that difference score is significantly affected by condition, I then created a model with condition and its associated slope removed[33] and compared this model to the previous model. The results showed the model including condition had a better fit than the model with this variable removed (log likelihood: -2712 versus -2729, ($\chi 2(2) = 33.9$, $p < .001$). This confirms that difference score is significantly affected by condition.

In particular, the results show that the difference score in Condition 2 ($M = 65.9$) is significantly higher than the difference score in Condition 1 ($M = -19.6$). These differences are shown in Figure 7 and suggest a post-stimulus increase in the length of the longest look-to-target in Condition 2, but not Condition 1. In this way, the results of the longest looks analysis support the findings of the proportional looks analysis and show that only the body orientation and pointing gesture was effective in directing participants attention to the target and away from the competitor. Neither analysis provides any evidence that participants are sensitive to body orientation alone.

[32] Formula: lmer(diffscore ~ condition + (1 + condition | participant), data=exp1.long.looks)
[33] Formula: model.3 = lmer(diffscore ~ age + (1 | participant), data=exp1.long.looks)

Figure 14: Difference score (ms) by condition, longest looks analysis, Experiment 2. Error bars show standard error



4.4.4 Additional Analyses

In the interests of thoroughness I also ran an analysis based on reaction time. This produced comparable results. There was insufficient data to run an analysis based on shift probability. Further details of both these analysis can be found in Appendix 4.

4.5 Discussion

This experiment tested the hypothesis that 2- and 3-year-olds would follow a gesture composed of two elements (i.e. body orientation and pointing) more than a gesture composed of just one element (i.e. body orientation). The predictions were that participants would (a) follow both types of gesture and (b) follow the body orientation and pointing gesture more than the body orientation only gesture. The first part of this prediction was not borne out: Participants followed the body orientation and pointing gesture but not the body orientation only gesture. The second part of the prediction was borne out: Participants followed the body orientation and pointing gesture more than they followed the body orientation only gesture. However, because there is no evidence that participants were sensitive to the body orientation gesture, their use of the body orientation and pointing gesture could be attributed to their sensitivity to the pointing component only. On this basis, I should not conclude that participants were more sensitive to the two-element gesture than the one-element gesture, but rather that they were more sensitive to pointing than body orientation. No developmental differences were found. The significance of these findings is discussed in more detail below.

Finding that 2- and 3-year-olds do not follow the body orientation gesture is unexpected as body orientation is a reliable indication of gaze direction and previous studies have shown that children younger than those tested here are sensitive to gaze direction. In particular, they both follow gaze direction and reliably use speakers' gaze direction to inform their referential choices and to map novel labels to unfamiliar objects (see Section 1.2.2 and 1.2.3). What, then, explains the apparent failure of the older children in the current study? One explanation for this is that the participants in the current study were not able to use body orientation to infer gaze direction. This has implications for how much children are able to learn from observing speech acts that are not directed to them. This is discussed in more detail below.

As discussed above, one theory of language acquisition posits that children can apply the same reasoning they use to infer a speaker's communicative intention when they are in a triadic interaction with the speaker to triadic interactions that they 'oversee', i.e. that they are not involved in (Tomasello, 2003). There is empirical evidence in support of this claim: From the age of 2;1 children can learn novel nouns through observation and from the age of 2;6 children can learn novel verbs through observation (Akhtar, Jipson, & Callanan, 2001). In the case of nouns, this finding is robust even when children are engaged in a potentially distracting task or when the label is presented in a non-declarative statement (Akhtar, 2005). In these studies 2-year-olds observed an interaction between two adults during which a novel label was used in reference to a novel object. No eye contact was made with the children during this interaction and no other attention was paid to them. At test, the participants reliably selected the labelled object as the intended referent of the novel label. These findings suggest that from the age of 2 children are adept at learning the meanings of new words from observation. In these studies, however, the speaker labelled the object (e.g. 'I'm going to show you the *toma*') before removing the target object from a box and handing it to the other adult participant. As such, the child observing the interaction did not have to rely on the speaker's gaze direction to infer her communicative intentions. The results of the present study suggest that children may not be so readily able to learn from observation if they have to use a speaker's gaze direction as a cue to reference and their eyes are not visible.

This has implications especially for cultures in which young children are not addressed directly and much of their language learning therefore comes from learning from conversations between others (de León, 2011; Küntay, Nakamura, & Ateş Şen, 2014; Lieven & Stoll, 2013; Mastin & Vogt, 2016). In this context, it is not possible to know exactly how much of the observed interactions occur when one or more of the participants' eyes are not visible to the child. Given the prevalence of baby-backing (where babies and infants are carried on the backs on their caregivers), however, it is likely that at least one participant's eyes are often hidden from view. If the current findings are accurate, backed babies may not be able to infer the gaze direction of the caregiver. This would suggest that word learning should be impaired in these conditions. This does not seem to be the case: These children learn language and learn from observed interactions. Indeed, Mastin and Vogt (2016) show that in rural Mozambique, where baby-backing is extremely prevalent, children's observation of interactions which they are not directly involved in is positively correlated with vocabulary size at 25 months of age. This finding therefore suggests that these children can learn through observation and are likely to be able to infer gaze direction even when the eyes are obscured. Why, then, did the 2- and 3-year-olds in the present study fail to do so? One possibility is that Western children have developed different monitoring skills compared with children who are expected to be keen observers (Akhtar, 2005). It is therefore possible that Western children do not infer gaze direction on the basis of body orientation and therefore cannot use body orientation to make referential choices. If so, this would suggest their ability to learn from overseen interactions may be limited. This conclusion may, however, be premature. This is discussed in more detail below.

An alternative explanation for the current findings is that 2- to 3-year-old children can make this inference, but that the participants in the present study did not do so for some reason. One potential reason is that they believed the looking action to be incidental rather than a deliberate signal of communicative intention. In particular, it has been shown that from the age of 14 months, infants can distinguish between intentional and accidental actions (Carpenter, Akhtar, & Tomasello, 1998). In the present experiment, as the actor looked towards the target she leant back, shifted her weight onto her back foot and moved her head and upper torso in the direction of the target. It is possible that participants interpreted this movement not as an intentional act pertaining to the labelling episode and produced for the benefit of the listener, but rather as an incidental act in which the actor was simply repositioning herself for her own comfort. As pointing is an inherently intentional act, used for the benefit of others, it is unlikely that participants would have made the same mistake in the body orientation and pointing condition. In this way, the results of this experiment could suggest that young children can distinguish between incidental and intentional actions, as well as accidental and intentional actions (see also Tomasello & Barton, 1994; Tomasello et al., 1996).

The suggestion that participants read the body orientation-only gesture as incidental rather than communicative or referential is supported by the fact that in the body orientation only condition participants tended to look at the speaker post stimulus (see Figure 11). One explanation for this is that they did not read the gesture as providing referential information and were seeking more information from the speaker. This corresponds to the finding that from the age of 18 months infants actively seek out referential cues from the speaker (Baldwin, 1991, 1993; Baldwin & Moses, 2001). One potential explanation for my participants' failure to follow the gesture in the body orientation only condition may be that they did not read the behaviour as a referential cue: Children do not make referential connections if they believe none were intended.

As discussed with regard to the null findings in Experiment 1, another explanation for participants' lack of attention to the body orientation cue is that both the target and the competitor belonged to the same category, e.g. two knife guards. As such, the cue was not disambiguating in the usual sense. Given that the same was true in the body orientation and pointing condition, the question then is why did participants attend to this cue? One possibility is that pointing is such a robust cue that children use it even when they are not engaged in making referential decisions. Given that pointing is regularly used to direct attention outside of the context of reference assignment this seems a likely explanation.

A final potential explanation for lack of significant results pertaining to body orientation is a mistake that was made in the preparation of the stimulus. Specifically, in both conditions the actor appears to speak while gesturing to the target, but no sound can be heard. This alone could explain why the participants looked at the speaker. It is important to note, however, that this lip movement was present in both conditions and does not appear to have affected the participants' behaviour in Condition 2. This provides additional evidence for the robustness of pointing as a tool for directing attention.

A surprising outcome from both Experiment 1 and Experiment 2 is participants' lack of attention to referential cues (specifically number marking and body orientation). As discussed in Section 1.4, previous studies have shown that at some points in development learners follow cues when they act in concert with each other in situations where they do not follow these same cues when they are presented in isolation. Experiment 3 therefore explores 2- and 3-year-olds use of these cues both in combination and in isolation to see if a similar pattern can be found.

Chapter 5 Experiment 3: Gesture and Grammatical Number

5.1 Aims and Predictions

As discussed in Section 1.4, there is evidence that although children are able to follow single cues, they show a greater tendency to follow cues that are acting in concert with each other. This is evident both in terms of their visual attention and in word learning. Furthermore, as discussed in Section 1.4.3, there is evidence that even if children do not follow cues in isolation they do follow those cues when they act together. Specifically, Hollich et al. (2000) found that 12-month-olds followed neither perceptual salience nor gaze direction when the two conflicted, but did follow the combination of those cues. Given the results of Experiment 1 and Experiment 2, it is this phenomenon that Experiment 3 considers: Even though 2- and 3-year-olds follow neither nominal number marking (Experiment 1) nor body orientation (Experiment 2) in isolation, do they follow the combination of these cues? I test this by comparing children's ability to follow these cues when they point to the same referent (congruent condition) and when they point to different referents (incongruent condition). As was the case in Experiments 1 and 2, I consider the issue developmentally by testing 2- and 3-year-olds. The reason for this is that if the results of Experiments 1 and 2 are not replicated, and the results of Experiment 3 show sensitivity to these cues in isolation, there may be developmental differences in terms of which cue (social or grammatical) children attend to when they conflict. In particular, as discussed previously, it is possible that across the course of development children shift from a reliance on social cues to a reliance on grammatical cues. It is therefore possible that if participants are sensitive to these cues in isolation, in the incongruent condition we will see older participants follow the number marking and younger participants follow the speaker's body orientation.

My final research questions is therefore as follows:

RQ3: Do 2- and 3-year-olds follow the combination of body orientation and nominal-number-marking more than they follow either cue in isolation. Does this change across development?

As noted above, this is tested by creating a context in which the cues either point to the same referent (the congruent condition) or to different referents (the incongruent condition). As described below, a combination of the stimuli used for Experiments 1 and 2 is used for Experiment 3. This is because Experiment 3 specifically compares the cues as they were presented in Experiments 1 and 2. Consequently, it continues to compare these cues when the referents belong to the same category and differ only in colour (e.g. one yellow knife guard versus 14 purple knife guards).

The main prediction that participants will pay more attention to a group of potential referents on congruent plural trials (when it is signalled by both plural number marking and body orientation) than on either incongruent plural trials (when it is signalled by plural number marking only) or incongruent singular trials (when it is signalled by body orientation only). Likewise, participants should pay more attention to the individual on congruent singular trials (when it is signalled by both singular number marking and body orientation) than on incongruent singular trials (when it is signalled by singular number marking only) or incongruent plural trials (when it is signalled by body orientation only). Given the results of Experiments 1 and 2, the prediction is that participants will pay little attention to either cue in the incongruent condition. If they do, however, it is predicted that older participants will follow grammatical number and younger participants will follow body orientation. The main variables of interest are therefore number, condition and age.

5.2 Methodology

5.2.1 Participants

Thirty-four 2- (n = 19, 10 girls) and 3-year-olds (n = 15, 4 girls) participated in this experiment (*M* = 2;11, Range = 2;0-3;11), none of whom had taken part in Experiment 1 or Experiment 2. All participants came from monolingual English-speaking homes (n = 31) or homes where English was the dominant language (n = 3). Participants were recruited and compensated as in Experiments 1 and 2. The experiment was approved by the University of Edinburgh Philosophy, Psychology and Language Sciences Ethics Committee. Eight additional children (four 2-year-olds and four 3-year-olds) participated in the experiment, but were excluded from analysis for the following reasons: restlessness (n = 4), experimenter error (n = 2) and trial loss (n = 2, see section 5.3 for details).

5.2.2 Design

The experiment used a 2x2 factorial design. The first variable was whether the gesture and the number marking were in agreement (congruent condition) or disagreement (incongruent condition). The second variable was number (singular or plural). These variables were fully crossed and varied within subjects. Four versions of the experiment were created to counterbalance 2 variables for each label: (a) target position (left or right) and (b) the object-label pairing.[34] The colour of the displays were balanced as in Experiments 1 and 2. Direction of turn was yoked to labels such that half of the labels always occurred with a clockwise turn and half always occurred with an anti-clockwise turn. Counter-balancing lists are reproduced in Appendix 3. Participants were assigned to lists pseudo-randomly such that equal numbers of participants completed each one.[35]

As before, there were two different types of trial: experimental trials in which unfamiliar objects were presented with novel labels and control trials in which familiar objects were presented with their conventional labels. All participants completed 16 experimental trials and 4 control trials.

5.2.3 Materials

Audio stimuli

The audio-stimuli were the stimuli used in Experiment 2 and Experiment 1, Condition 1 (i.e. singular and plural sentences containing number marking on the noun only, e.g. ‘I found the *zoot/s*’.)

Visual Stimuli

The visual stimuli were the displays from Experiment 1 (i.e. one display showing a group of objects and the other display showing a single object) against the videos from Experiment 2, Condition 1 (i.e. with the speaker turning to face one of the displays).

5.2.4 Trial procedure

In all trials two images were displayed: one of a single object and one of a group of objects, both presented against a white background. The single object and the multiple objects were always the same category and different colours as illustrated in Figure 15.

[34] The object-label pairing was varied for unfamiliar objects and labels only; familiar objects were always presented with their conventional labels.
[35] Following participant loss, the final proportions were as follows: 8/34 completed List A; 8/34 completed List B, 11/34 completed List C; and 7/34 completed List D.

Figure 15: Visual stimuli in situ, Experiment 3

The images were presented against the videos used in Experiment 2, Condition 1 in which the actor turns and appears to face either the display on the left or the display on the right. In the congruent condition the visual stimuli matched the audio stimuli: The speaker looked at the individual object and produced a singular sentence (congruent singular trials) or the speaker looked at the group of objects and produced a plural sentence (congruent plural trials). In the incongruent condition, the audio stimuli did not match the visual stimuli: The speaker either looked at the individual object and produced a plural sentence (incongruent plural trials) or looked at the group of objects and produced a singular sentence (incongruent singular trials). The overall trial timings were as in Experiments 1 and 2 and detailed timings are found in Figure 16.

Figure 16: Trial progression, Experiment 3

Time (ms)	**Audio**	**Visual**			
		Actor	**Object**	**Condition 1 (congruent)**	**Condition 2 (incongruent)**
0		Facing camera	Absent		
1000	Boing!	Facing camera	Present		
2080	Gasp!	Facing camera	Present		
3160	‘Let’s see!’	Turning to wall	Present		
4000		Turning to wall	Present		
5000	‘Look!’	Facing wall	Present		
6000		Facing wall	Present		
7280	‘I found...	Facing wall	Present		
8000	... the *zoot*!’	Starts gesturing	Present		
9000		Gesturing	Present		
10000		Gesturing	Present		
11000		Gesturing	Present		
12000		Stops gesturing	Present		
13000	Boing!	Turning to camera	Absent		
14000		Turning to camera	Absent		
15000		Turning to camera	Absent		
16000		Facing camera	Absent		

5.2.5 Session Procedure

As Experiments 1 and 2, with the exception that two participants preferred to watch the videos without their parent.

5.3 Coding

The coding procedure was the same as in Experiments 1 and 2.[36] Again, a second coder independently complete data sets from eight participants (24% of the useable data). Four were randomly selected from the 2-year-old participants and four from the 3-year-old participants. Agreement between the coders was high. When all four coding categories were used, the mean agreement between coders was 85% (*SD* = 4.28, Cohen’s κ = 0.77). This indicates a substantial degree of agreement (Landis & Koch, 1977). When the ‘speaker’ and ‘elsewhere’ codes were collapsed, mean agreement

[36] In this experiment, two participants spontaneously pointed at the displays. In doing so, their arms obscured or partially obscured their eyes for 35-45 ms. On these occasions, direction of point, rather than direction of gaze, was recorded.

increased to 90% (*SD* = 3.20, Cohen's κ = 0.82). This indicates a 'near-perfect' degree of agreement (Landis & Koch, 1977). When the coders disagreed, the first coder's decision was used. Videos were coded such that coders did not know which experimental condition any given trial belonged to.

The data were then filtered such that any trial on which greater than or equal to 20% of frames were coded 'elsewhere' was excluded from analysis.[37]

5.4 Analysis and Results

5.4.1 Eye-movements

As in Experiments 1 and 2, the eye-movement data suggest participants were engaged and on-task (see Figure 17 - Figure 20). As in Experiment 1, participants showed an initial preference for the individual object when the stimuli first appeared on screen (1000ms).[38] This attention to the individual was followed by a sustained period of looks-to-speaker as she spoke and turned to face the wall (c.2250-4750ms). This is followed by a second peak in looks-to-speaker following the onset of the gesture and test sentences (c.9000ms). Following this peak, looks-to-group and looks-to-individual are relatively level for singular congruent trials and plural incongruent trials. This suggests that participants did not look more at the individual when the speaker looked it, regardless of whether this gesture was accompanied by singular number marking (i.e. in the congruent singular condition) or when it was accompanied by contradictory plural number marking (i.e. in the incongruent plural condition). In the congruent plural condition and the incongruent singular condition there appears to be a post-stimulus peak in looks-to-group. This could suggest that participants looked more at the group when the speaker looked at the group both when it was supported by plural number marking (congruent plural condition) and when it was contradicted by singular number marking (incongruent singular condition). However, in both cases the 'peak' is broadly comparable to the looks-to-group in the pre-stimulus window. This suggests that participant's tendency to look at the group in these conditions was not the result of either

[37] Comparable results were found when these data were included. One hundred and ninety-eight out of 720 trials were excluded in this way. This equates to approximately 28% of the useable data collected. This left 522 trials available for analysis. Following this loss 2 participants failed to contribute at least one trial per condition. As reported in 5.2.1, their data (20 trials) were excluded from further analysis. In addition, one participant ended the experiment before completing the final trial. This left 501 trials available for analysis. Of these, 101 were control trials and 400 were experimental trials. The following analyses were conducted with the experimental trials only.

[38] Note that this was not case in Experiment 2 as the displays showed either two groups or two individuals.

the number marking or gesture cues. The following sections explore the extent to which these impressions are borne out by statistical analysis.

Figure 17: Looks to group (red), individual (green), speaker (blue) and elsewhere (purple), Experiment 3, Congruent Condition, singular trials

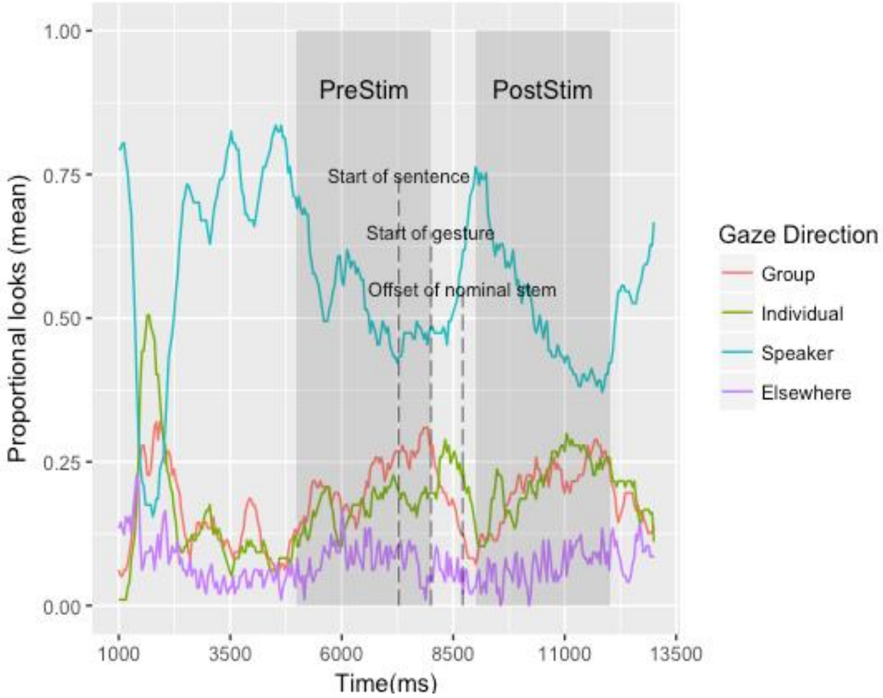


Figure 18: Looks to group (red), individual (green), speaker (blue) and elsewhere (purple), Experiment 3, Congruent Condition, plural trials

Figure 19: Looks to group (red), individual (green), speaker (blue) and elsewhere (purple), Experiment 3, Incongruent Condition, singular trials

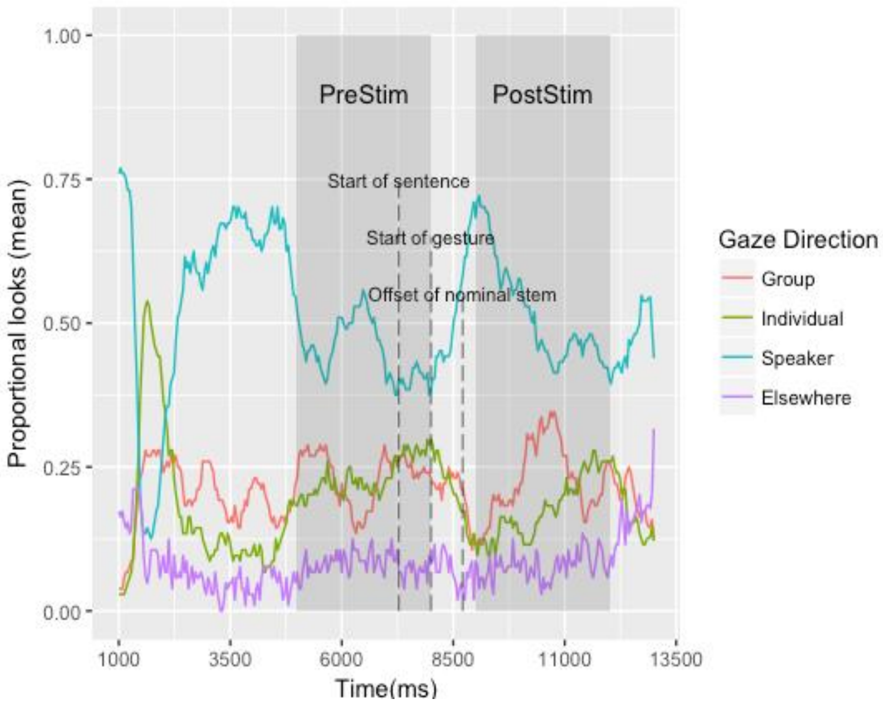
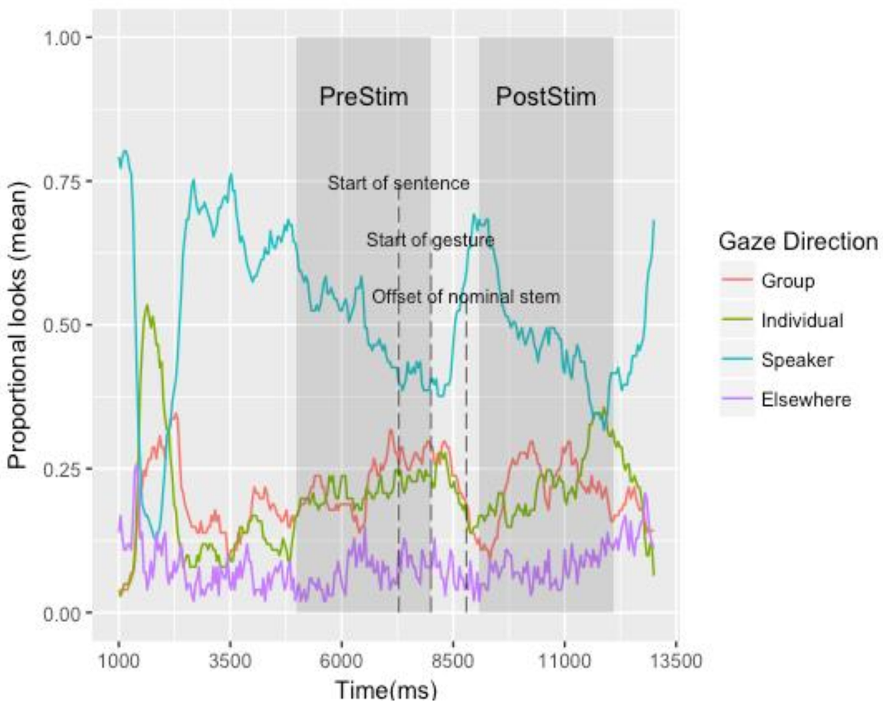


Figure 20: Looks to group (red), individual (green), speaker (blue) and elsewhere (purple), Experiment 3, Incongruent Condition, plural trials



5.4.2 Proportional looks

Analysis

As in Experiment 1, the dependent variable in this analysis is calculated by subtracting the proportion of looks-to-group in the pre-stimulus window from the proportion of looks-to-group in the post-stimulus window to give a difference score.[39] A positive difference score indicates a post-stimulus increase in looks-to-group. A negative difference score indicates a post-stimulus increase in looks-to-individual. If the results of Experiment 1 and Experiment 2 are robust, and neither cue is effective in directing attention in isolation, the post-stimulus difference score in the incongruent condition should not differ significantly from zero for either singular or plural trials. If my hypothesis is correct, and cues acting in concert are more effective in directing attention than cues acting in isolation, difference scores in the congruent condition should be significantly greater than zero for plural trials and significantly less than zero for singular trials.

Following Experiment 2, the duration of the analysis windows is set at 3000ms. Again, as in Experiment 2, the pre-stimulus window extends back from the onset of the gesture and therefore runs between 4000ms and 7000ms. During this period participants have been exposed to neither stimulus. As in Experiment 1, the post-stimulus window extends from 300ms after the offset of the nominal stem (i.e. 300ms after | in 'I found the *zoot* | !' or 'I found the *zoot* | s!'). During this period participants have been exposed to both stimuli. The start of the post-stimulus window varies across items and the mean is 7711ms for singular trials and 7793 for plural trials. The average location is marked in grey in Figure 17 to Figure 20.

Results

The descriptive statistics are given in Table 13. All values used to calculate the means are within 2.5 standard deviations of the grand mean.

[39]To calculate a difference score, participants therefore had to have looked at either the target or the competitor at least once in both the pre- and post-stimulus windows. This was not the case on 77 out of 400 trials. These trials were excluded from analysis. Following this loss, 2 participants failed to contribute data to all four conditions. Their data was excluded from this analysis. This analysis was therefore run with the data from remaining 32 participants only (94%). This equates to 316 trials (79%).

Table 13: Difference score means by condition, number, age and allomorph, proportional looks analysis, Experiment 3

	Mean (SD)	Mean (SD) by age		Mean (SD) by allomorph	
		2yo	3yo	/s/	/z/
Congruent Singular	-0.17 (0.36)	-0.22 (0.37)	-0.10 (0.36)	NA	NA
Congruent Plural	-0.01 (0.32)	-0.13 (0.31)	0.13 (0.28)	-0.01 (0.41)	0.02 (0.47)
Incongruent Singular	0.07 (0.36)	0.10 (0.34)	0.03 (0.40)	NA	NA
Incongruent Plural	-0.05 (0.31)	-0.05 (0.33)	-0.05 (0.31)	0.00 (0.38)	-0.09 (0.42)

As in Experiments 1 and 2, I assessed the impact of the independent variables on difference score by using the lme4 package (Bates et al., 2012) for R (R Core Team, 2019) to perform a linear mixed effects analysis. The model was created by entering number, condition and age (months) as fixed effects. The categorical variables condition and number and were made numeric by assigning the value -0.5 to trials in Condition 1 (congruent condition) and 0.5 to trials in Condition 2 (incongruent condition) and -0.5 to singular trials and 0.5 to plural trials. Age was centred by taking each participant's age in months and subtracting the mean. The variables were added to the model with interaction terms. As random effects the model had an intercept for participant and by-participant random slopes for number and condition.[40] The results of this model are shown in Table 14.

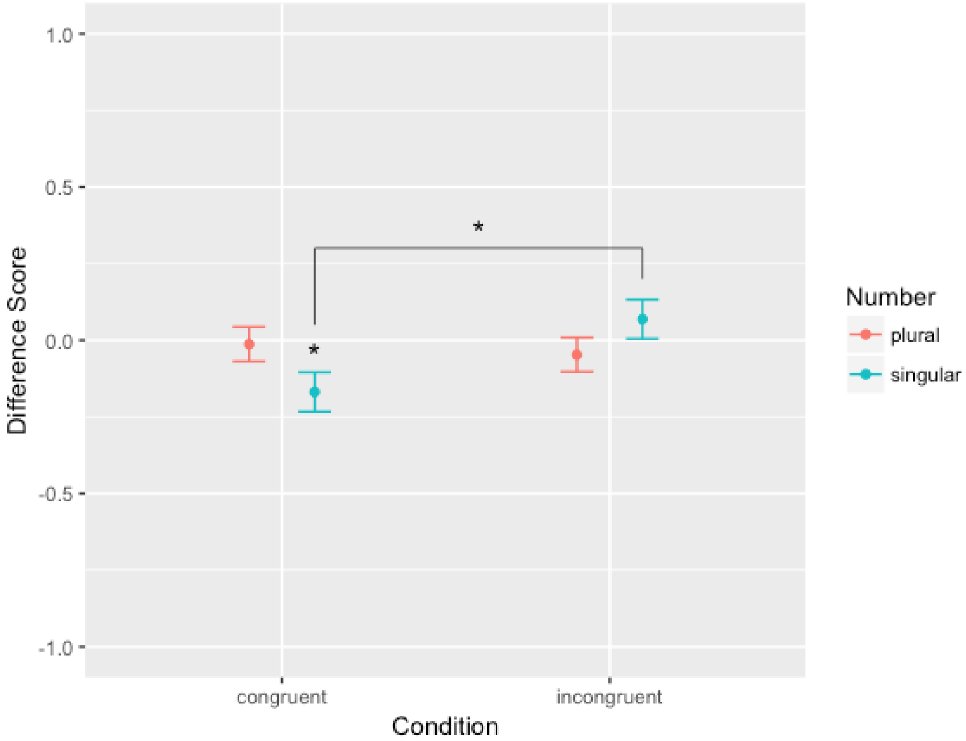
Table 14: Results of a linear mixed effects model examining the effect of number, condition and age on difference score, proportional looks analysis, Experiment 3

Factor	**Estimate**	**Std. Error**	**df**	***t***	***P***	
Intercept	-0.0153	0.0279	109.16	-0.547	0.58523	
Number	0.0157	0.0575	51.12	0.273	0.78627	
Condition	-0.0828	0.0569	60.49	-1.455	0.15079	
Age	0.0047	0.0040	110.85	1.169	0.24492	
Number:Condition	0.2880	0.1107	298.35	2.601	0.00976	**
Number:Age	0.0099	0.0083	50.51	1.192	0.23877	
Condition:Age	0.0135	0.0082	58.97	1.652	0.10391	
Number:Condition:Age	0.0027	0.0159	296.31	0.170	0.86529	

[40] Formula: model.full = lmer(diffscore ~ number * condition * age + (1 + number + condition | participant), data=exp3.prop.looks)

The results show a significant interaction between number and condition. This interaction is illustrated in Figure 21.

Figure 21: Difference score by condition and number, proportional looks analysis, Experiment 3



As shown in Figure 21, on congruent trials the difference score for singular trials was significantly lower than zero (M = -0.17, $t(31) = -2.63$, $p < .01$). None of the other scores differed significantly from zero. This suggests that on congruent singular trials (i.e. trials where the speaker looked at the individual object and used singular number marking), there was a significant increase in participants' tendency to look at the individual. This was not the case when the speaker looked at the individual but used plural number marking (i.e. on incongruent plural trials) or when she used singular number marking but looked at the group (i.e. on incongruent singular trials). Furthermore, the mean difference sore on singular trials in the congruent condition (M = -0.17) was significantly lower than the mean difference score on singular trials in the incongruent condition (M = 0.07, $t(31) = -2.32$, $p < .05$). These findings suggest that, for singular but not plural trials, participants were sensitive to the cues when they acted in concert, but not when they acted in isolation.

Although the model did not indicate an effect of age, I check for developmental differences by looking at 2- and 3-year-olds separately. In particular, I ran one sample t-test comparing the mean difference score to zero for both age groups on singular trials in the congruent condition. The results showed that while the mean difference score for 2-year-olds was significantly lower than zero (M = -.22, $t(17) = - -2.56$, $p < .05$), the score for 3-year-olds did not differ significantly from zero (M = -0.10, t (13) = -1.04, p = 0.32). This suggests that the 2-year-olds were more likely to look at the individual object after hearing singular number marking and seeing the speaker face the individual object, but the 3-year-olds were not.

Overall, the results of the proportional looks analysis suggest that 2-year-old participants show sensitivity to the combination of singular number marking and gaze direction. This is not the case for plural number marking. To check whether this was the case for labels taking both the /s/ allomorph and labels taking the /z/ allomorph, I created an additional model with allomorph, condition and age as fixed effects. Condition was made numeric as before. Allomorph was made numeric by assigning 0.5 to trials where the label was inflected /s/ and -0.5 to trials where the label was inflected /z/. Age was centred as before. As random effects the model had an intercept for participant and a by-participant random slope for allomorph and condition. This model was run with plural trials only and the results showed a significant interaction between age, condition and allomorph (see Table 15).

Table 15: Results of a linear mixed effects model examining the effect of age, condition and allomorph on difference score, proportional looks analysis, Experiment 3

Factor	Estimate	Std. Error	df	*T*	*p*	
Intercept	-0.0058	0.0398	30.29	-0.144	0.8861	
Age	0.0107	0.0057	29.66	1.881	0.0698	.
Condition	0.0781	0.0755	38.05	1.035	0.3071	
Allomorph	-0.0074	0.0751	28.79	-0.099	0.9220	
Age:Condition	0.0152	0.0108	36.71	1.415	0.1655	
Age:Allomorph	0.0059	0.0107	28.10	0.555	0.5832	
Condition:Allomorph	-0.0829	0.1382	126.15	-0.600	0.5497	
Age:Condition:Allomorph	-0.0432	0.0196	113.12	-2.206	0.0294	*

To explore this interaction further I consider the congruent and incongruent conditions separately. Beginning with the congruent condition, I created a mixed effects model looking at the effect age and allomorph on difference score. As fixed effects the model had age (months) and allomorph and as a random effect it has an intercept for participant and a by-participant random slope for the effect of allomorph.[41] This model was fit with the plural data from the congruent condition and showed a significant main effect of age (t = 2.27, p <. 05) and no significant interaction between age and allomorph. To investigate the effect of age I ran three t-tests comparing 2- and 3-year-olds difference score to zero and to each other. None of the tests reached significance. To explore this effect further I divided the congruent data into 4 groups based on age in months. As before, this was done to assess the possibility that the developmental differences found involved only the youngest 2-year-olds or oldest 3-year-olds (rather than 2- and 3-year-olds in general) Age group 1 contained the 8 youngest participants (24-28 months), Age group 2 the next 8 youngest (29-35 months), Age group 3 the next 8 youngest (35-41 months) and Age group 4 the 8 oldest (42-46 months).[42] The descriptive statistics are given in Table 16.

Table 16: Difference score means by age group, congruent condition (plural trials), proportional looks analysis Experiment 3

	Age Group 1	**Age group 2**	**Age Group 3**	**Age Group4**
Mean (sd)	-0.22 (0.47)	0.05 (0.38)	0.10 (0.35)	0.14 (0.52)

Planned comparisons showed the difference score for Age group 2 (M = 0.05) was significantly higher than the difference score for Age group 1 (M = -0.22). As the difference scores for Age group 3 (M = 0.10) and Age group 4 (M = 0.14) were higher than the difference score for Age group 2, and the difference score for Age group 2 was significantly higher than the difference score for Age group 1, it can be inferred that the difference scores for Age group 3 and Age group 4 were also significantly higher than the difference score for Age group 1. Recall that a positive difference score reflects the increase in the proportion of looks to group. The fact that Age groups 2-4 produce positive difference scores, while Age group 1 produces a negative difference score indicates that older participants' tendency to look at the group increased post-stimulus while the youngest participants' tendency to look at the group decreases. As these difference scores are calculated for plural trials in the congruent condition only, it can be concluded that older participants were more guided by the combination of plural number marking and the speaker looking towards the group than younger participants.

[41] Formula: lmer(diffscore ~ age * allomorph + (1 + allomorph| participant), data=Cond1)
[42] When participants were tied on age in months they were assigned to groups based on age in months and days.

To explore the relationship between age and allomorph in the incongruent condition, I fit the model described above with the incongruent plural data. The results showed a significant interaction between age and allomorph (t = 2.13, p < .05). To explore this interaction, I began by comparing 2- and 3-year-olds. The descriptive statistics are given in Table 17.

Table 17: Mean difference score by age and allomorph on plural trials, incongruent condition, Experiment 3

	2-year-olds		3-year-olds	
	/s/	/z/	/s/	/z/
Mean (sd)	-0.14 (0.52)	0.06 (0.48)	0.08 (0.41)	-0.18 (0.41)

I then ran t-tests comparing 2- and 3-year-old's difference scores to zero on both /s/ trials and /z/ trials and also t-tests comparing 2- and 3-year-olds difference scores on /s/ trials and their scores on /z/ trials. Although none of the tests reached significance, the tests involving 3-year-olds use of the /z/ allomorph were much closer to doing so than the others. In particular, the mean difference score for 3-year-olds on trials with the /z/ allomorph was -0.18. This was near-significantly different from zero (t(16) = -1.8, p = .09). Furthermore, this score was near-significantly different from the difference score for 3-year-olds on trials with the /s/ allomorph (M = 0.08, t(33) = 1.85, p = .07) and from the difference score for 2-year-olds on trials with the /z/ allomorph (M = 0.06, t(38) = 1.73, p = .09). Further analyses involving either breaking the sample down into smaller age groups or treating age as a continuous variable yielded neither significant nor near-significant results.[43] We therefore conclude that, despite not reaching significance, 3-year-olds' response to the /z/ allomorph was the developmental difference the model picked up on. Recall that a negative difference score indicates an increase in looks-to-individual. These results therefore suggest that on plural trials with the /z/ allomorph 3-year-olds' attention to the individual increased. This is the opposite of what we would expect and supports the overall finding that this analysis provides no evidence that participants were sensitive to plural marking (even in combination with a gesture towards the group). In the following section we see whether this is also the case for the longest looks analysis.

[43] We did not run a follow-up analysis with control items because both familiar labels (spoons and chairs) took the {z} allomorph.

5.4.3 Longest looks

Analysis

The dependent variable in this analysis is calculated by subtracting the longest look-to-individual from the longest look-to-group in both the pre-stimulus and post-stimulus windows to create two difference scores. The pre- and post-stimulus windows were located as before. The pre-stimulus difference score was then subtracted from the post-stimulus difference score to yield a third difference score. A positive difference score therefore indicates an increase in attention to group and a negative difference score indicates an increase in attention to the individual. If participants are not sensitive to either number marking nor gaze direction alone (as the results of Experiments 1 and 2 suggest), then in the incongruent condition difference scores on both singular and plural trials should not differ from zero. If participants are sensitive to the combination of plural number marking and looks-to-group, then in the congruent condition difference score on plural trials should be significantly greater than zero. If they are sensitive to the combination of singular number marking and look-to-individual, then in the congruent condition difference score on singular trials should be significantly lower than zero.

Results

The descriptive statistics are shown in Table 18. Difference scores that were 2.5 or more standard deviations from the mean were excluded.[44]

Table 18: Difference score means (ms) by condition, age and allomorph, longest looks analysis, Experiment 3

	Mean (SD)	**Mean (SD) by age**		**Mean (SD) by allomorph**	
		2yo	3yo	/s/	/z/
Congruent Singular	-9.12 (114.8)	-27.54 (137.9)	14.22 (74.7)	NA	NA
Congruent Plural	-0.39 (86.5)	-26.67 (93.06)	32.89 (66.3)	18.00 (116.7)	-0.67 (128.0)
Incongruent Singular	-0.22 (87.4)	17.33 (80.8)	-22.44 (93.2)	NA	NA
Incongruent Plural	1.33 (123.8)	12.88 (148.2)	-13.07 (86.4)	4.14 (125.6)	-21.08 (143.1)

[44] Six out of 400 trials were excluded as a result. The following analysis was run with the remaining 394 trials only (99%). A follow-up analysis in which these values were included yielded comparable results. Following this data loss all participants contributed data to all four conditions.

As before, I assessed the impact of the independent variables on difference score by using the lme4 package (Bates et al., 2012) for R (R Core Team, 2019) to perform a linear mixed effects analysis. The model was created by entering number, condition and age (months) as fixed effects. The categorical variables were made numeric and age was centred as before. The variables were added to the model with interaction terms. As random effects the model had an intercept for participant and by-participant random slopes for number and condition.[45] The results of this model are given in Table 19.

Table 19: Results of a linear model examining the effect of number, condition and age on difference score, longest looks analysis, Experiment 3

Factor	Estimate	Std. Error	df	*t*	*P*
Intercept	-1.2232	8.6162	27.87	-0.142	0.888
Number	12.6449	16.7692	50.39	0.754	0.454
Condition	-3.4636	16.2833	63.87	-0.213	0.832
Age	0.2931	1.2060	26.02	0.243	0.810
Number:Condition	14.5609	30.5570	349.17	0.866	0.391
Number:Age	2.0304	2.3436	47.23	0.866	0.391
Condition:Age	3.2950	2.2724	59.60	1.450	0.152
Number:Condition:Age	-4.9990	4.2512	343.82	-1.176	0.240

As shown, there are no significant main effects or interactions. In particular, there is no interaction between number and condition. This suggests that difference scores were comparable for singular and plural trials in both conditions. As such, there is no evidence that participants looked longer at the group after hearing plural sentences, or at the individual after hearing singular sentences. This was the case both when the number marking was accompanied by a supporting body orientation gesture and when it was accompanied by a contradictory one. This analysis therefore supports Experiments 1 and 2 in showing that participants are not sensitive to either number marking or body orientation when these cues occur in isolation. Furthermore, it does not support the results of the proportional looks analysis in that it provides no evidence of sensitivity to the combination of singular number marking and body orientation towards the individual. Follow up tests confirmed this was the case for both 2- and 3-year-olds.

To test whether participants were any more sensitive to the labels inflected /s/ than those inflected /z/, I created a third model which included allomorph as a fixed effect.

[45] Formula:model.full = lmer(diffscore ~ number * condition * age + (1 + number + condition | participant), data=exp3.long.looks)

This model took allomorph, age and condition as fixed effects. Allomorph was made numeric and age was centred as before. As random effects the model had an intercept for participant and by-participant random slopes for allomorph and condition.[46] This model was run with plural trials only. As shown in Table 20, there were no significant main effects or interactions. This suggests that in both conditions participants looked no longer at the group of items after hearing labels inflected with /s/ than labels inflected with /z/.

Table 20: Results of a linear mixed effects model examining the effect of age, condition and allomorph on difference score, longest looks analysis, Experiment 3

Factor	Estimate	Std. Error	df	*t*	*P*
Intercept	5.5253	13.0861	32.53	0.422	0.676
Age	1.4194	1.8192	29.88	0.780	0.441
Condition	6.7966	23.6111	40.09	0.288	0.775
Allomorph	2.8614	19.9443	168.19	0.143	0.886
Age:Condition	1.1653	3.2613	36.39	0.357	0.723
Age:Allomorph	-0.5564	2.7211	163.91	-0.204	0.838
Condition:Allomorph	-23.5895	39.9410	169.86	-0.591	0.556
Age:Condition:Allomorph	-7.9367	5.4475	165.39	-1.457	0.147

Additional Analyses

Due to data loss it was not possible to run a reaction time or shift probability analysis with this data. Details are given in Appendix 4.

5.5 Discussion

This experiment showed that although 2- and 3-year-olds follow neither grammatical number marking (Experiment 1) nor body orientation (Experiment 2), 2-year-olds, but not 3-year-olds, do follow these cues when they are presented together. The fact that this pattern was shown in the proportional looks analysis, but not the longest looks analysis, suggests the former may be a more appropriate measure for children of this age in this type of task. More specifically, the results showed that when a speaker turned and looked at an individual object and labelled that object with a novel noun marked as singular, 2-year-olds looked longer at the individual object than at the group of objects. This was not the case for plural trials: When a speaker turned and looked at

[46] Formula: lmer(diffscore ~ age * condition * allomorph + (1 + allomorph + condition | participant), data=plu)

the group and labelled those objects with a novel noun marked as plural, 2-year-olds did not look longer at the group. For the 3-year-olds this tendency increased with age.

Recall that the congruent condition presented the nominal number marking cue and the body orientation cue in alignment. The previous experiments showed that 2- and 3-year-olds use neither of these cues in isolation. Following the cues on the congruent trials therefore suggests children are sensitive to the effect of cue quantity: They follow cues in combination when they do not follow them in isolation. The developmental differences observed therefore suggest that for 3-year-olds the effect of cue quantity increased across development on plural trials, while for 2-year-olds the effect of cue quantity decreased across development on singular trials. This suggests that the effect of cue quantity does not straightforwardly increase or decrease across development. As such, these developmental differences are not revealing about the origins of the observed preference for multiple cues (see section 1.2.2). Alternative explanations for these developmental differences are given below.

Looking in more detail at the behaviour of the two-year-olds, what might explain the asymmetry between singular and plural trials? One explanation is that the single object was more perceptually salient than the group of objects. Two-year-olds may have looked at the single, large picture to gain more information about the category of the objects. If this is the case, there would have been three cues pointing to the individual object as the intended referent of the novel label on singular trials (i.e. perceptual salience, body orientation, and number marking) and only two cues pointing to the group of objects as the intended referent on plural trials (i.e. body orientation and number marking). The fact that this pattern was found only for 2-year-olds, and not 3-year-olds, suggests that the role of perceptual salience and/or cue quantity may be limited to the earlier stages of development (although extending later than shown by previous studies, e.g. Hollich et al. (2000)). It is not possible to tell at this point whether this finding should be attributed to the role of cue quantity or perceptual saliency. This is a logical problem (one which I did not appreciate at the outset of this research): If participants respond to x + y more than x it is not clear whether that response should be attributed to the role of x + y per se or to the effect of y only. This point is returned to in Chapter 6. Nonetheless, the fact that older children might not be affected by the perceptual saliency of the referents in this task raises an interesting question about the window for learners' use of perceptual saliency as a referential cue. In particular, it suggests that children's use of perceptual saliency cues may change between the ages of 2 and 3 years.

On congruent plural trials (i.e. when the speaker looked at the group and produced a plural label), neither 2-year-olds nor 3-year-olds paid more attention to the group than the individual. For 3-year-olds, however, the tendency to do so improved with age. One explanation for this is that older children do not require the additional support of saliency cues and/or cue quantity. This, however, would not explain why they failed on singular trials. Taken jointly with the fact that younger children followed the combination of body orientation and nominal singular number marking, an alternative explanation is that sensitivity to singular number marking emerges before sensitivity to plural number marking, and that older children therefore performed better on plural trials because they were more sensitive to plural number marking. This is surprising as in English plural number marking on nouns is overt (e.g. I found the *zoot***s)** and singular number marking on nouns is null (e.g. I found the *zoot****Ø***). As such, we would therefore expect the sensitivity to the plural to emerge before sensitivity to the singular. Indeed, there is existing evidence that this is the case (for an overview see Davies, Xu Rattanasone, Schembri, et al., 2019). This finding is therefore intriguing, and suggests more work should be done regarding developmental differences in the emergence of sensitivity to singular and plural number marking. I return to this point in Chapter 6 in my discussion of directions for future research.

An alternative explanation for the developmental differences found in this experiment might be that the 2-year-olds were more sensitive to the grammatical and social cues in combination than the 3-year-olds. Given, though, that the ability to follow social and grammatical cues has to be learned across development this does not seem likely.

A final potential explanation for these differences is sample size. The proportional looks analysis (where the developmental difference were found) was conducted with eighteen 2-year-olds and fourteen 3-year-olds. Although this is a small difference it is possible that I did not collect enough data from 3-year-olds to find a significant effect. This is supported by the fact that the results for 3-year-olds were in the expected direction.

Returning to the asymmetry between singular and plural trials, the perceptual salience of the singular referent is not the only explanation: There is a logical reason for why this might have occurred. In particular, while the combination of the individual and the group could be construed as a group, only the individual could be construed as an individual. That is, on hearing a plural, participants may have looked at both the individual and the group, but on hearing a singular they should, logically, have looked only at the individual. This could explain participants' improved performance on singular trials. This problem is inherent in using objects of the same category for both the singular and plural referent and one solution would be to consider plural trials only. One factor that speaks against this interpretation, however, is the developmental differences observed: Why would younger participants behave more logically than older participants? Another is the looking behaviour of the participants in the post stimulus windows of the plural trials: If they were interpreting the plural noun as referring to the group and individual combined we should observe an increase in looks to both displays. As Figure 18 and Figure 20 show, however, this was not the case. Instead, they show an increase in looks to speaker. This suggests that although this interpretation was a logical possibility, it was not one that the participants made.

The main finding of this experiment was that 2-year-old participants followed cues in combination that they ignored in isolation. This could support the claim that multiple cues facilitate attention. In this way, these findings add to the literature discussed in Section 1.4 showing a facilitative effect of multiple cues on attention. Furthermore, the behaviour of the 3-year-olds suggests that either the effect of perceptual salience and/or the effect of cue quantity declines across development. Of course, these two factors could be related as the effect of cue quantity could ultimately be attributable to perceptual saliency.

In this way, these experiments provide new evidence that learners follow cues in combination even though they ignore those cues in isolation. This was found only when the cues were of different types. What might account for this? One possibility is that by presenting the cues in combination the same information was presented across two modalities. This proposal is discussed in more detail below.

In the examples presented in Experiments 1 and 2, the information was presented in a single modality. In particular, comparing body orientation with body orientation and pointing involves single versus multiple types of visual stimuli. Similarly, comparing nominal with nominal and verbal number marking involves single versus multiple types of auditory stimuli. Experiment 3, however, compares information presented in one modality (i.e. auditory cues in Experiment 1 and visual cues in Experiment 2) with information presented simultaneously in two modalities (visual and auditory cues in Experiment 3). Some researchers have suggested that information presented across two modalities might be particularly salient. For example, according to the Intersensory Redundancy Hypothesis, information that is presented across multiple modalities becomes foregrounded and stands out from the surrounding environment (Bahrick, Lickliter, & Flom, 2004). Furthermore, it is argued that ‘learning amodal properties [such as object-label pairings] is facilitated in multi-model stimulation’ (Bahrick et al., 2004, p101). Evidence in support of this comes from the facilitative effect of object movement in referential decision-making tasks (Houston-Price et al., 2006; Werker, Cohen, Lloyd, & Casasola, 1998). For example, Gogate and Bahrick (1998) tested the ability of 7-month-olds to form object-label mappings in three conditions: moving synchronous, when the object moved in time with the vocalic label; moving asynchronous, when the object moved out of time with the vocalic label; and still, when the object did not move during the presentation of the vocalic label. The results showed participants only formed a new object-label pairing in the moving synchronous condition. This suggests that presenting the cues across two modalities (and making sure the cues were in time) facilitated referential understanding.

Rader and Zukow-Goldring (2012) provide similar evidence with regard to the acquisition of novel object-label pairings and synchronous versus asynchronous gestures. In this experiment, 9- to 15-month-old infants were shown object-label pairings in two conditions. The first compared a dynamic synchronous gesture with no gesture and the other compared a dynamic synchronous gesture and a dynamic asynchronous gesture. The results showed that infants associated labels with referents more when they were presented with a dynamic synchronous gesture than with either a dynamic asynchronous gesture or with no gesture. Again this suggests that information presented synchronously across two modalities helps young learners acquire new object-label mappings. The findings of the current experiment may therefore be explained by the fact that not only were there two cues, but that the two cues were provided synchronously in two modalities. Previous evidence in support of the Intersensory Redundancy Hypothesis has come primarily from pre-verbal children (for a review see Bahrick et al., 2004). This experiment therefore provides new evidence that multi-modal input is also facilitative in cue following for older learners.

Chapter 6 General Discussion

The aim of this thesis was to shed light on the reference problem and make progress in understanding how children know what novel labels mean. In particular, it explored the role of cue quantity and asked whether social and grammatical cues that were comprised of multiple sub-cues were more effective in directing attention than those which comprised only one cue.

Experiment 1 considered this issue with regard to grammatical cues and tested the hypothesis that 2- and 3-year-olds will follow the combination of nominal and verbal number marking more than nominal number marking alone. Contrary to previous studies, the results showed that participants followed neither cue. This was attributed to the fact that, unlike in previous experiments, all the potential referents belonged to the same category. Participants were therefore faced with the task of working out ‘what goes with what’. I argue that this is more similar to the task children are faced with if they are to follow (and use) grammatical number in the real world. As such, the results of previous studies may overestimate children’s ability in these regards. It is therefore possible that grammatical number is a less useful tool for solving the reference problem than previously believed.

Experiment 2 considered social cues and asked whether children’s attention was more guided by the combination of body orientation and pointing than by body orientation alone. The results were surprising and showed that while 2- and 3-year-olds followed body orientation and pointing they did not follow body orientation alone. Two possible explanations for this finding are that (a) children could not infer gaze direction on the basis of body orientation or (b) they did not read the body orientation cue as informative.

Experiment 3 provided the only evidence we have seen in support of the effect of cue quantity. In particular, the behaviour of the 2-year-olds suggested that when perceptual, social and grammatical cues are in alignment (as they were on singular trials in the congruent condition) they are sufficient for directing attention. When all three cues are not acting in concert this is not the case. This suggests either that the role of perceptual saliency varies between the ages of 2 and 3 years of age, that the role of cue quantity varies between these ages, or both. This is discussed in more detail below.

One of the main aims of the current work was to look for developmental differences as a means of teasing apart different explanations of the observed behaviour. This was not possible in Experiments 1 and 2 as no developmental differences were found. In

Experiment 3, however, developmental differences were found. In particular, Experiment 3 showed that although 2- and 3-year-olds followed neither grammatical number marking (Experiment 1) nor body orientation (Experiment 2), 2-year-olds did attend to these cues when they were presented together. That is, when a speaker faced an individual object and labelled that object with singular nominal number marking, 2-year-olds looked longer at the individual object than at the group of objects. One explanation for this is that in Experiments 1 and 2 the stimuli were presented in only one modality (auditory and visual respectively). In Experiment 3 however, they were presented across two modalities (visual and auditory). As discussed, it is possible that this highlighted the information and led the participants to attend to it. It is not immediately clear why this was the case for singular items, but not plural items, i.e. why did participants not look longer at the group than the individual when the speaker looked at the group and labelled it using plural morphology? This could be explained by the fact that the individual object is more salient than the group and that once participants engaged in the task of referent assignment (which they did not do in Experiment 1) they were more drawn to the individual than the group. This perceptual salience, coupled with the social and grammatical cues, may therefore have facilitated participants' ability to map the singular label to the singular referent. That this was the case only for 2-year-olds, and not 3-year-olds, suggests that the role of perceptual salience and/or cue quantity changes between 2 and 3 years.

The design of Experiment 3 was very similar to a recently published study by Paquette-Smith and Johnson (Paquette-Smith & Johnson, 2016). As with the present experiment, this experiment used a preferential looking design in which participants were shown pictures of a group of objects and an individual object. In this case, the objects belonged to different categories. Participants then heard sentences containing number marked determiners, nouns and verbs (e.g. '**This is a** nice *zurpel*' or '**These are** nice *zurpel***s**'). In line with previous findings in this area (e.g. (Kouider et al., 2006), the results showed that even 24- to 25-month-olds looked longer at the group after hearing sentences containing plural number marking and at the individual after hearing sentences containing singular number marking. Furthermore, they selected the object(s) picked out by the number marking as the intended referent(s) of the novel label at test. As stated previously, the difference between the present results and those of Paquette-Smith and Johnson (2016) is attributed to the former using referents belonging to different categories and including a repetition of the number marking (see Section 3.5). In the second experiment, participants heard and saw the same stimuli except that a speaker either looked at the matching object (i.e. at the individual object on singular trials and the group of objects on plural trials, Convergent Condition) or the non-matching object (i.e. at the group of objects on singular trials and the individual object on plural trials, Divergent Condition). In this case, the speaker faced forward and looked either down and left or down and right. In keeping with the present results, evidence of learning was found only in the Convergent Condition; in the Divergent Condition participants were at chance. The authors interpret this as showing that participants are already sensitive to grammatical number to the extent that it interferes with learning through gaze direction.

A surprising outcome of my experiments is the null results with regard to grammatical number in Experiment 1, body orientation in Experiment 2 and for plural trials in the congruent condition in Experiment 3. This leads me to question whether my analyses were powerful enough to detect positive results As has already been noted, the exclusion criteria used reduced the number of data points available and in doing so reduced the power of the analyses. As such, the null results obtained might be because the analyses were underpowered. Using the simr and pwr packages for R (R Core Team, 2019) I calculated the power of the key analyses that produced null results. These are listed in Table 21. As shown, the power is very low in all cases. This suggests that these analyses were underpowered and that a far larger sample size would be required to elicit a significant result. However, many previous studies looking at similar phenomena have found significant results with comparable sample sizes to mine (see Table 22 below). This indicates that the analyses that produced null results were measuring something different from these studies. It should therefore be concluded that these analyses provide no evidence of sensitivity to number marking, body orientation or the combination of the two.

Table 21: Power by experiment and analysis

Exp.	Analysis	Test	Power
1	Proportional looks	Interaction between Number, Condition and Age	17%
1	Longest looks	Interaction between Number, Condition and Age	9%
2	Proportional looks	Difference score to zero in Condition 1	15%
2	Longest looks	Difference score to zero in Condition 1	17%
3	Proportional looks	Difference score to zero on plural trials in congruent condition (2yos)	17%
3	Longest looks	Interaction between Number, Condition and Age	21%

Another relevant issue here is whether the number of participants sampled was sufficient, especially when looking for developmental differences. Experiment 1 tested

33 participants, Experiment 2 tested 41 and Experiment 3 tested 34. As shown in Table 22 these sample sizes are broadly comparable to other studies in this area. In studies looking for developmental differences, though, the number of participants in each age group is generally higher and the age range more narrow (e.g. comparing 18- and 24-month-olds rather than 2- and 3-year-olds). The number of experimental trials I conduct (N = 16), however, is on the higher end of the spectrum. For this reason I believe the overall sample size to be appropriate even when looking for developmental differences.

Table 22: Sample size in representative studies using the preferential looking paradigm

Study	Exp.	Age of Participants	No. Participants	No. trials	Topic
Briganti and Cohen (2011)	1	14- and 18-month-olds	48 (24 14mos and 24 18mos)	4	Pointing
Hollich et al. (2000)	1	12-, 19- and 24-month-olds	96 (32 12mos, 32 19mos and 32 24mos)	2	Perceptual salience and gaze direction
Hollich et al. (2000)	2	12-, 19- and 24-month-olds	36 (12 12mos, 12 19mos and 12 24mos)	2	Perceptual salience and gaze direction
Houston-Price et al. (2006)	1	15-month-olds	27	12	Perceptual salience and gaze direction
Houston-Price et al. (2006)	2	15-month-olds	30	12	Perceptual salience and gaze direction
Houston-Price et al. (2006)	3	15-month-olds	30	12	Perceptual salience and gaze direction
Houston-Price et al. (2006)	4	15-month-olds	32	12	Perceptual salience and gaze direction
Jolly and Plunkett (2008)	1	2-year-olds and 30-month-olds	72 (39 2yos and 33 30mos)	4	Number marking
Kouider et al. (2006)	1	2-year-olds	14	10	Number marking
Kouider et al. (2006)	2	20-month-olds	16	10	Number marking
Kouider et al. (2006)	3	2-year-olds	16	10	Number marking
Kouider et al. (2006)	4	20-month-olds	12	10	Number marking
Lukyanenko and Fisher (2016)	1	3-year-olds	32	16	Number marking
Lukyanenko and Fisher (2016)	2	30-month-olds	32	16	Number marking
Nappa et al. (2009)	1	3-, 4- and 5-year-olds	39 (11 3yos, 12 4yos and 16 5yos)	7	Gaze direction, conceptual and linguistic cues
Nappa et al. (2009)	2	3-, 4- and 5-year-olds	53 (12 3yos, 25 4yos and 16 5yos)	7	Gaze direction, conceptual and linguistic cues
Paquette-Smith and Johnson (2016)	1	22- to 25-month-olds	17	4	Number marking and gaze direction
Paquette-Smith and Johnson (2016)	2	24- and 25-month-olds	17	4	Number marking and gaze direction
Pruden et al. (2006)	1	10-month-olds	77	4	Perceptual salience and gaze direction
Pruden et al. (2006)	2	10-month-olds	52	4	Perceptual salience and gaze direction
Schafer and Plunkett (1998)	1	12- to 17-month-olds	29	12	Word learning

One of the strengths of the present work is that the main results are supported by multiple analyses. For example, Experiment 1 showed that participants were not sensitive to number marking. This was the case in both the proportional looks and longest looks analyses. Similarly, Experiment 2 showed that participants were sensitive to body orientation and pointing but not to body orientation alone. This was found in both the proportional looks and longest looks analyses and also in a supplementary analysis based on reaction time (see Appendix 4). Experiment 3, however, showed significant results only in the proportional looks analysis. It was not supported by the longest looks analysis. Having converging results from multiple analyses allows for confident interpretation of the results. It also allows researchers to assess which type of analysis is most effective in analysing this kind of data.

To elaborate, Table 23 shows the four different types of analyses I pursued (proportional looks, longest looks, reaction time and shift probability). Of these, I was only able to complete the proportional looks and longest looks analyses for all three experiments (completed analyses are in black and abandoned analyses are in red). For the reaction time and shift probability analyses there was generally not enough data. This was partly due to the criterion that all participants contribute data to all conditions and partly due to the calculation of the dependent variable (see analysis sections for details). This suggests that one would need to collect more data to run analyses based on reaction time and shift probability than to run analyses based on proportional or longest looks. For this reason the latter are preferable to the former, at least with participants of this age.

Table 23: Samples size by analysis type, Experiments 1-3. Red text indicates abandoned analyses.

Exp.	Analysis	Ppts included	Ppts sampled	% ppts included	Trials included	Trials available[47]	% trials included
1	Proportional looks	32	33	97	354	402	88
1	Longest looks	32	33	97	384	402	96
1	Reaction time	24	33	73	196	402	49
1	Shift probability	4	33	12	68	402	17
2	Proportional looks	40	41	98	391	432	91 Eh
2	Longest looks	41	41	100	422	432	98
2	Reaction time	40	41	98	312	432	72
2	Shift probability	18	41	44	151	432	35
3	Proportional looks	32	34	94	316	400	79
3	Longest looks	34	34	100	394	400	99
3	Reaction time	18	34	53	142	400	36
3	Shift probability	0	34	0	0	400	0

There were various factors involved in deciding which analyses to perform. In some cases the decision was simple: In Experiments 1 and 3 preparing the data for the shift probability analysis left what was clearly too little data (i.e. no data at all in Experiment 3 and 68 trials across 4 participants in Experiment 1). For Experiment 2 it was less clear cut, but with fewer than half the participants sampled and fewer than half the trials they completed remaining I decided not to pursue this analysis. With regard to reaction time, in Experiment 3 only around half the participants (53%) and around a third of trials (36%) were available for analysis. In Experiment 1 there were a reasonable number of participants remaining (73%), but only half the trials (49%). To decide whether or not to include a particular analysis I used the criteria that at least 50% of the data (both in terms of participants and trials) had to be included. The analyses marked in red in Table 23 failed to meet these criteria.

One limitation of the present work is that the experiments presented here do not fit neatly with the existing literature. In particular, all three experiments used potential referents belonging to the same category and differing only in colour. As such, the referential problem was not the classic 'what category of object does this novel label

[47] This is the number of trials completed on which participants were looking 'elsewhere' on fewer than 20% of frames.

refer to', but rather 'what does this speaker mean to refer to when she uses this novel label'. These processes are referred to as resolving word reference and resolving speaker reference respectively. According to one dominant account of word learning, however, word reference and speaker reference are essentially the same thing (Tomasello, 2001, 2003). To elaborate, on this account word learning does not involve the learning of an abstract mapping between label and referent, but rather involves working out what a speaker is using a label to refer to in a particular context. In short, words do not refer, speakers do. In this way, the distinction between word reference and speaker difference may not be so great as it first appears. As such, this distinction may not be the reason the present experiments did not replicate results from previous work.

A second factor that separates the present work from much of the existing literature is that it does not include an explicit test of participants' understanding of the novel label. As discussed in Chapter 2, many studies of this type include a training phase, in which the cues are produced and novel labels introduced, followed by a test phase, in which no cues are produced and understanding of the previously introduced novel label is assessed (see for example Hollich et al., 2000; Houston-Price et al., 2006; Jolly & Plunkett, 2008; Pruden et al., 2006). This is not always the case, however. Some previous studies used visual attention as a proxy for referential understanding and assume that increased visual attention to a target following a label implies the formation of a referential link between the two (e.g. Kouider et al., 2006). Indeed, this is the assumption on which the preferential looking paradigm is predicated (see Chapter 2). The latter approach was adopted for three reasons.

The first reason was incidental: The design of Experiment 1 was based on the design used by Kouider et al. (2006) in which no test phase was used. Although criticised by other researchers (e.g. Jolly & Plunkett, 2008; Paquette-Smith & Johnson, 2016), Kouider and colleagues discuss their results in terms of the participants' referential use of grammatical number marking.

The second reason was practical: Including a test phase greatly increases the length of each trial and, cumulatively, increases the length of the experiment. I was keen to employ a within-subjects design for the key variables (see Schafer & Plunkett, 1998) and pilot runs during the design phase suggested children of the age tested here would not tolerate experiments of greater than 15 minutes in total. I therefore prioritised having a within-subjects design over having a test phase.

The third reason was conceptual: I believed increased visual attention to an object on hearing a label did indeed imply a referential link between the two. As discussed in Chapter 2, evidence for this comes from the fact that in some studies that include a test phase, participants' behaviour in the training phase is accurately reflected in the test phase. That is, increased visual attention to the target during training is reflected in increased visual attention to the target at test. Since conducting the experiments, however, it has come to my attention that performance in the training phase does not always predict performance in the test phase. That is, the object the participants look at longest during training is not always the object they select as the intended referent at test. For example, in studies that manipulate the perceptual saliency of the potential referents, participants tend to look longer at the salient object during training, even if they do not select that object as the intended referent at test (Hollich et al., 2000; Houston-Price et al., 2006; Moore et al., 1999). For studies that manipulate perceptual saliency this is perhaps unsurprising; one of the referents is selected to be or made to be more interesting and attention-grabbing than the other.

However, this is also the case in studies where perceptual saliency is not manipulated and the potential referents are assumed to be approximately balanced in terms of their saliency. For example, in the training phase of Briganti and Cohen (2011)'s study both 14- and 18-month-olds looked longer at the object the experimenter pointed to than the object she did not point to. However, only the 18-month-olds selected the object she pointed to as the intended referent of the novel label at test. This indicates that cue following during training does not entail cue use at test, even when perceptual saliency is not a factor. Furthermore, the results of Paquette-Smith and Johnson (2016) show that even when behaviour is superficially comparable in the training and test phases, one does not necessarily predict the other. To elaborate, in the training phase of this study participants looked longer at the group of objects when they heard plural number marking and longer at the individual object when they heard singular number marking. They then went on to select the object or group of objects that the number marking picked out during training at test. This, therefore, appears to be an example of behaviour in the training phase predicting behaviour in the test phase. However, as the authors point out, the two were not correlated. This suggests that although the overall behaviour of the participants was comparable in training and test, this was not the case for individual participants. In other words, the amount of time an individual looked at the target during training did not predict whether they selected that item at test. As a result, I retract my previous statement that 'increased visual attention reliably correlates with referential understanding' and acknowledge that if the results presented here are to be applied to word learning they would need to be replicated in experiments that include a test phase. The extra time this would add to the experiment could be offset by varying the main independent variable between subjects, rather than within subjects. As discussed in more detail below, this represents an avenue for future research.

At this point, the results pertain to visual attention to a target, not referential understanding. My expectation is that, despite that patterns discussed above, increased visual attention to a target will indeed facilitate acquisition of the label used for that target. That is, although increased visual attention to the target in training does not predict increased visual attention to the target at test, an effect of increased visual attention during training (as a result of cue quantity or otherwise) could be observed in measures such as speed of referent selection at test or retention of the novel label post test. Further experiments are needed to determine whether this is the case. The results presented here therefore represent a first step in determining how factors like cue quantity affect word learning.

An additional limitation of the present studies is that they are not fully able to address the effect of cue quantity on participants' increased visual attention to a target. This is because if children respond more to a sentence or a gesture comprising x + y than to a sentence or gesture comprising x, one cannot rule out that this increase in attention is caused by y rather than the combination of x + y. To illustrate, even if my experiment had shown that children's referential choices were more guided by nominal and verbal number marking than nominal number marking alone, it would not be clear whether this was due to the combination of nominal and verbal marking (i.e. cue quantity) or due to the influence of verbal number marking only. The same is true for the two element versus one element gesture comparison; if children are more attentive to body orientation and pointing than body orientation alone, one cannot rule of that this increase in attention is caused by the addition of pointing rather than the combination of body orientation and pointing. This is a logical problem that I did not anticipate when designing the experiments. One way to unravel this problem is by comparing x and y in isolation and comparing each to x + y. This represents a clear direction for future research.

In particular, Experiments 1 and 2 could be usefully expanded by including a third condition testing the second part of the multiple cue in isolation. In Experiment 2, this would be a pointing only condition. In Section 1.3 I suggested that pointing gestures are unlikely to occur naturalistically without accompanying eye-gaze. My design circumvents this problem by having the actor face away from the participant when she produces the gesture. This modification could be neatly incorporated into the updated methodology described above: three between subject conditions (body orientation, pointing, and body orientation plus pointing) composed of trials including test phases.

The problem is not so neatly solved with regard to grammatical number. In particular, although it is possible to produce sentences with verbal number marking only by using zero plurals (e.g. ‘There are the sheep’) these are known to pose a particular problem for learners (see Section 1.2). In particular, we know that 30-month-olds expect number agreement between subjects and verbs. As such, participants may perform poorly on zero plural trials not because they are not sensitive to verbal number marking, but because they perceive a mismatch between subject and verb. An alternative option would be to move from grammatical cues to semantic cues and look at the effect of stacking adjectives. For example, do participants look longer at a target that is signalled by multiple adjectives (e.g. ‘Look at the blue sparkly one!’) than by a single adjective (e.g. ‘Look at the blue one!’ or ‘Look at the sparkly one!’).

One of the main outstanding issue is whether the finding that 2- and 3-year-olds do not follow grammatical number when the referents belong to the same category is a genuine finding or an experimental construct. To investigate this, I plan to run follow-up experiments using the same basic methodology, with potential referents belonging to the same category and to different categories. If I find that children do follow and use grammatical number in the latter but not the former I will have conclusive evidence that the findings presented here are not the result of methodological issues, but rather are informative with regard to children’s ability to follow grammatical number. As discussed above, these experiments would differ from those presented here as they will include a test phase and the main independent variable would be varied between subjects.

Additionally with regard to grammatical number marking, the results of Experiment 3 provide new evidence that for English-speaking children sensitivity to the singular might emerge before sensitivity to the plural. As noted in Section 5.5, this contradicts previous research suggesting sensitivity to the plural emerges before sensitivity to the singular (Davidson et al, 2019). The reason given for the earlier acquisition of the plural is that in English singular number marking is null and plural number marking is overt. One avenue for exploring this further would therefore be to compare English-speaking children’s acquisition of ‘is’ and ‘are’, where both singular and plural marking are overt. In addition, research could compare learners of English (in which singular marking in null and plural marking overt) with learners of a language where this is not the case and both singular and plural number marking are overt.

Another outstanding issue is 2- and 3-year-olds' failure to follow the body orientation cue in Experiment 2. As noted above, there are two possible explanations for this: (a) children could not infer gaze direction on the basis of body orientation or (b) they did not read the body orientation cues as informative. Another direction for further research is to distinguish between these possibilities and determine whether 2- and 3-year-olds can make this kind of inference. This could be done by making the intentionality of the action clearer. One way of doing this would be to present the action along with a sentence explaining the movement, e.g. 'Now I can see!' or 'That's better', as the actor turns to face the target. If learners still fail to follow the body orientation cue under these conditions that will constitute evidence that they do not follow body orientation in tasks that have a referential component. Under these same conditions I found that participants did respond appropriately to the body orientation and pointing cue. This suggests that pointing is a robust cue and is a particularly effective tool for directing young children's attention.

An additional direction for further research concerns the developmental window for perceptual salience and/or cue quantity. In particular, the results of Experiment 3 suggest that while 2-year-olds are guided by perceptual salience and/or cue quantity, 3-year-olds are not. To explore this further, I suggest running experiments with 2- and 3-year-olds in which perceptual saliency or cue quantity are the only cues to reference. The prediction would be that only the younger participants would show evidence of word learning. Crucially, testing perceptual salience and cue-quantity in a context where they are not confounded would allow us to disentangle them as potential explanation for the behaviour observed in Experiment 3.

A final direction for further research that emerges from Experiment 3 and Paquette-Smith and Johnson (2016) is to run this kind of experiment with older children. In particular, it would be enlightening to run an experiment with potential referents of different categories and the speaker facing forward but with older children. In Paquette-Smith and Johnson (2016)'s study the participants had just turned two and were just starting to be able to use grammatical number as a cue to reference. It is possible that if a similar study were run with older children they would show evidence of learning in the divergent conditions and pay more attention to the grammatically cued display than the socially cued display (see Nappa et al., 2009).

In summary, the findings presented here show that some cues are more effective than others when it comes to directing visual attention. In particular, it shows that pointing remains a particularly robust cue throughout development and suggests that body orientation may be a less useful cue. Furthermore, it suggests that grammatical number may be less useful than has previously been believed. Finally, it suggests that either perceptual saliency or the combination of perceptual saliency, social cues and grammatical cues improves visual attention. Which of these is the better explanation is a question for future research.

Appendices

Appendix 1: Written description of the experiment given to parents

Thank you for volunteering to help with my research

Please read the following information carefully and let me know if you have any questions.

You and your child have been invited to take part in a study I am running as part of my PhD research. The topic of the study is word learning and I am looking at whether children's ideas about what words mean changes depending on the sentences and gestures speakers use when those words are first introduced.

If you agree to take part in this study you and your child will sit together and watch a number of short videos on a computer screen. While your child watches the videos you will be asked to close your eyes and listen to some music. Underneath the screen, mostly hidden from view, is a video camera that will be recording your child as they watch the videos. I will then use these recordings to see what children are paying attention to at different times, and whether this changes depending on what the actor in the videos says and does.

The videos are designed to be fun and engaging for young children, and families typically enjoy their visit to the lab. However, if for any reason your child becomes unhappy and does not wish to continue you are free to stop at any time. After watching the videos, you will be asked to complete a short questionnaire and be given a £2.50 voucher for a local coffee shop to thank you for your participation. Your child will be given their choice of a selection of sticker books and thanked for their help.

If you would like to take part then please complete the attached consent form.

If you have any questions at any time please feel free to ask.

Appendix 2: Informed consent form

PARENT/GUARDIAN CONSENT FORM
I have read the attached information and **I would like** my child to take part in the described study.

I understand that my child will be audio- and video-recorded during the study. I also understand that all data collected will be protected and kept confidential, and our anonymity will be maintained in all published and written work resulting from this the study.

I understand that there are no risks associated with this task and that my child and I are free to withdraw from the study at any point, with no need to justify our decision, even after this form has been signed.

I understand that this project has received ethical approval from the University of Edinburgh

PARENT/GUARDIAN'S NAME:……………………………………………………………………….

CHILD'S NAME: ……………………………………………………………………………….

CHILD'S GENDER:………………………………………………………………………………….

CHILD'S DATE OF BIRTH: ……………………………………………………………………..

CONTACT DETAILS (please provide at least one contact):

E-mail:……………………………………………………………………………………………..

Telephone number:………………………………………………………………………………

DATE:……………………………………………………………………………...

SIGNATURE:…………………………………………………………………………..
(Parent / Guardian)

If you are willing to allow (parts of) the audio-visual recordings made of your child to be presented to other researchers, for example at an academic conference or in a lecture, please sign below. Please note that your child's name and other personal details will **not** be used in this context.

SIGNATURE:…………………………………………………………………………..
(Parent / Guardian)

If you are willing to allow still images taken from the audio-visual recording to be used in an academic context, for example published in an academic journal, or presented at a conference or a lecture, please sign below. Please note that your child's name and other personal details will **not** be used in this context.

SIGNATURE:…………………………………………………………………………..
(Parent / Guardian)

Appendix 3: Counterbalancing lists

Experiment 1: List A

No.	Cond.	Trial type	Label	Allo-morph	Object	Target Pos.	Target Colour	Target Type	Comp. Pos.	Comp. Colour	Comp. Type	Turn
1	1	exp.	feek		1	left	purple	individual	right	yellow	individual	CW
2	1	exp.	glaive		2	left	red	individual	right	green	individual	ACW
3	1	exp.	zoot		3	right	green	individual	left	red	individual	ACW
4	1	exp.	tulb		4	right	yellow	individual	left	purple	individual	CW
5	1	exp.	feeks	s	1	right	yellow	group	left	purple	group	CW
6	1	exp.	glaives	z	2	right	green	group	left	red	group	ACW
7	1	exp.	zoots	s	3	left	red	group	right	green	group	ACW
8	1	exp.	tulbs	z	4	left	purple	group	right	yellow	group	CW
9	1	exp.	feek		1	left	purple	individual	right	yellow	individual	CW
10	2	exp.	glaive		2	left	red	individual	right	green	individual	ACW
11	2	exp.	zoot		3	right	green	individual	left	red	individual	ACW
12	2	exp.	tulb		4	right	yellow	individual	left	purple	individual	CW
13	2	exp.	feeks	s	1	right	yellow	group	left	purple	group	CW
14	2	exp.	glaives	z	2	right	green	group	left	red	group	ACW
15	2	exp.	zoots	s	3	left	red	group	right	green	group	ACW
16	2	exp.	tulbs	z	4	left	purple	group	right	yellow	group	CW
17	1	control	chair		5	left	blue	individual	right	green	individual	CW
18	1	control	spoons		6	right	green	group	left	blue	group	ACW
19	2	control	spoon		6	left	blue	individual	right	green	individual	ACW
20	2	control	chairs		5	right	green	group	left	blue	group	CW

Experiment 1: List B

No.	Cond.	Trial type	Label	Allo-morph	Object	Target Pos.	Target Colour	Target Type	Comp. Pos.	Comp. Colour	Comp. Type	Turn Direction
1	1	exp.	feek		1	right	purple	individual	left	yellow	individual	CW
2	1	exp.	glaive		2	right	red	individual	left	green	individual	ACW
3	1	exp.	zoot		3	left	green	individual	right	red	individual	ACW
4	1	exp.	tulb		4	left	yellow	individual	right	purple	individual	CW
5	1	exp.	feeks	s	1	left	yellow	group	right	purple	group	CW
6	1	exp.	glaives	z	2	left	green	group	right	red	group	ACW
7	1	exp.	zoots	s	3	right	red	group	left	green	group	ACW
8	1	exp.	tulbs	z	4	right	purple	group	left	yellow	group	CW
9	1	exp.	feek		1	right	purple	individual	left	yellow	individual	CW
10	2	exp.	glaive		2	right	red	individual	left	green	individual	ACW
11	2	exp.	zoot		3	left	green	individual	right	red	individual	ACW
12	2	exp.	tulb		4	left	yellow	individual	right	purple	individual	CW
13	2	exp.	feeks	s	1	left	yellow	group	right	purple	group	CW
14	2	exp.	glaives	z	2	left	green	group	right	red	group	ACW
15	2	exp.	zoots	s	3	right	red	group	left	green	group	ACW
16	2	exp.	tulbs	z	4	right	purple	group	left	yellow	group	CW
17	1	control	chair		5	right	blue	individual	left	green	individual	CW
18	1	control	spoons		6	left	green	group	right	blue	group	ACW
19	2	control	spoon		6	right	blue	individual	left	green	individual	ACW
20	2	control	chairs		5	left	green	group	right	blue	group	CW

Experiment 1: List C

No.	Cond.	Trial type	Label	Allo-morph	Object	Target Pos.	Target Colour	Target Type	Comp. Pos.	Comp. Colour	Comp. Type	Turn Direction
1	1	exp.	feek		3	left	green	individual	right	red	individual	CW
2	1	exp.	glaive		4	left	yellow	individual	right	purple	individual	ACW-
3	1	exp.	zoot		1	right	purple	individual	left	yellow	individual	ACW -
4	1	exp.	tulb		2	right	red	individual	left	green	individual	CW
5	1	exp.	feeks	s	3	right	red	group	left	green	group	CW
6	1	exp.	glaives	z	4	right	purple	group	left	yellow	group	ACW -
7	1	exp.	zoots	s	1	left	yellow	group	right	purple	group	ACW -
8	1	exp.	tulbs	z	2	left	green	group	right	red	group	CW
9	1	exp.	feek		3	left	green	individual	right	red	individual	CW
10	2	exp.	glaive		4	left	yellow	individual	right	purple	individual	ACW -
11	2	exp.	zoot		1	right	purple	individual	left	yellow	individual	ACW -
12	2	exp.	tulb		2	right	red	individual	left	green	individual	CW
13	2	exp.	feeks	s	3	right	red	group	left	green	group	CW
14	2	exp.	glaives	z	4	right	purple	group	left	yellow	group	ACW -
15	2	exp.	zoots	s	1	left	yellow	group	right	purple	group	ACW -
16	2	exp.	tulbs	z	2	left	green	group	right	red	group	CW
17	1	control	chair		5	left	blue	individual	right	green	individual	CW
18	1	control	spoons		6	right	green	group	left	blue	group	ACW ACW
19	2	control	spoon		6	left	blue	individual	right	green	individual	ACW -
20	2	control	chairs		5	right	green	group	left	blue	group	CW

Experiment 1: List D

No.	Cond.	Trial type	Label	Allomorph	Object	Target Pos.	Target Colour	Target Type	Comp. Pos.	Comp. Colour	Comp. Type	Turn Direction
1	1	exp.	feek		3	right	green	individual	left	red	individual	CW
2	1	exp.	glaive		4	right	yellow	individual	left	purple	individual	ACW
3	1	exp.	zoot		1	left	purple	individual	right	yellow	individual	ACW -
4	1	exp.	tulb		2	left	red	individual	right	green	individual	CW
5	1	exp.	feeks	s	3	left	red	group	right	green	group	CW
6	1	exp.	glaives	z	4	left	purple	group	right	yellow	group	ACW -
7	1	exp.	zoots	s	1	right	yellow	group	left	purple	group	ACW -
8	1	exp.	tulbs	z	2	right	green	group	left	red	group	CW
9	1	exp.	feek		3	right	green	individual	left	red	individual	CW
10	2	exp.	glaive		4	right	yellow	individual	left	purple	individual	ACW -
11	2	exp.	zoot		1	left	purple	individual	right	yellow	individual	ACW -
12	2	exp.	tulb		2	left	red	individual	right	green	individual	CW
13	2	exp.	feeks	s	3	left	red	group	right	green	group	CW
14	2	exp.	glaives	z	4	left	purple	group	right	yellow	group	ACW -
15	2	exp.	zoots	s	1	right	yellow	group	left	purple	group	ACW -
16	2	exp.	tulbs	z	2	right	green	group	left	red	group	CW
17	1	contr	chair		5	right	blue	individual	left	green	individual	CW
18	1	contr	spoons		6	left	green	group	right	blue	group	ACW ACW
19	2	contr	spoon		6	right	blue	individual	left	green	individual	ACW -
20	2	contr	chairs		5	left	green	group	right	blue	group	CW

Experiment 2: List A

No	Cond.	Trial type	Number	Label	Allo-morph	Obj.	Target Pos.	Target Colour	Target Type	Comp Pos.	Comp Colour	Comp Type	Turn Direction
1	1	exp.	singular	feek		1	left	purple	ind.	right	yellow	group	CW
2	1	exp.	singular	glaive		2	left	red	ind.	right	green	group	ACW
3	1	exp.	singular	zoot		3	right	green	ind.	left	red	group	ACW
4	1	exp.	singular	tulb		4	right	yellow	ind.	left	purple	group	CW
5	1	exp.	plural	feeks	s	1	right	yellow	group	left	purple	ind.	CW
6	1	exp.	plural	glaives	z	2	right	green	group	left	red	ind.	ACW
7	1	exp.	plural	zoots	s	3	left	red	group	right	green	ind.	ACW
8	1	exp.	plural	tulbs	z	4	left	purple	group	right	yellow	ind.	CW
9	2	exp.	singular	feek		1	left	purple	ind.	right	yellow	group	CW
10	2	exp.	singular	glaive		2	left	red	ind.	right	green	group	ACW
11	2	exp.	singular	zoot		3	right	green	ind.	left	red	group	ACW
12	2	exp.	singular	tulb		4	right	yellow	ind.	left	purple	group	CW
13	2	exp.	plural	feeks	s	1	right	yellow	group	left	purple	ind.	CW
14	2	exp.	plural	glaives	z	2	right	green	group	left	red	ind.	ACW
15	2	exp.	plural	zoots	s	3	left	red	group	right	green	ind.	ACW
16	2	exp.	plural	tulbs	z	4	left	purple	group	right	green	ind.	CW
17	1	control	singular	chair		5	left	blue	ind. dual	right	green	group	CW
18	1	control	plural	spoons		6	right	green	group	left	blue	ind.	ACW
19	2	control	singular	spoon		6	left	blue	ind. dual	right	green	group	ACW
20	2	control	plural	chairs		5	right	green	group	left	blue	ind.	CW

Experiment 2: List B

No.	Cond.	Trial type	Number	Label	Allo-morph	Obj.	Target Pos.	Target Colour	Target Type	Comp. Pos.	Comp. Colour	Comp Type	Turn Direction
1	1	exp.	singular	feek		1	right	purple	ind.	left	yellow	group	CW
2	1	exp.	singular	glaive		2	right	red	ind.	left	green	group	ACW
3	1	exp.	singular	zoot		3	left	green	ind.	right	red	group	ACW
4	1	exp.	singular	tulb		4	left	yellow	ind.	right	purple	group	CW
5	1	exp.	plural	feeks	s	1	left	yellow	group	right	purple	ind.	CW
6	1	exp.	plural	glaives	z	2	left	green	group	right	red	ind.	ACW
7	1	exp.	plural	zoots	s	3	right	red	group	left	green	ind.	ACW
8	1	exp.	plural	tulbs	z	4	right	purple	group	left	yellow	ind.	CW
9	2	exp.	singular	feek		1	right	purple	ind.	left	yellow	group	CW
10	2	exp.	singular	glaive		2	right	red	ind.	left	green	group	ACW
11	2	exp.	singular	zoot		3	left	green	ind.	right	red	group	ACW
12	2	exp.	singular	tulb		4	left	yellow	ind.	right	purple	group	CW
13	2	exp.	plural	feeks	s	1	left	yellow	group	right	purple	ind.	CW
14	2	exp.	plural	glaives	z	2	left	green	group	right	red	ind.	ACW
15	2	exp.	plural	zoots	s	3	right	red	group	left	green	ind.	ACW
16	2	exp.	plural	tulbs	z	4	right	purple	group	left	green	ind.	CW
17	1	control	singular	chair		5	right	blue	ind. dual	left	green	group	CW
18	1	control	plural	spoons		6	left	green	group	right	blue	ind.	ACW
19	2	control	singular	spoon		6	right	blue	ind. dual	left	green	group	ACW
20	2	**dd**	plural	chairs		5	left	green	group	right	blue	ind.	CW

Experiment 2: List C

No.	Cond.	Trial type	Number	Label	Allo-morph	Obj.	Target Pos.	Target Colour	Target Type	Comp. Pos.	Comp. Colour	Comp Type	Turn Direction
1	1	exp.	singular	feek		3	left	green	ind.	right	red	group	CW
2	1	exp.	singular	glaive		4	left	yellow	ind.	right	purple	group	ACW
3	1	exp.	singular	zoot		2	right	purple	ind.	left	yellow	group	ACW
4	1	exp.	singular	tulb		1	right	red	ind.	left	green	group	CW
5	1	exp.	plural	feeks	s	3	right	red	group	left	green	ind.	CW
6	1	exp.	plural	glaive	z	4	right	purple	group	left	yellow	ind.	ACW
7	1	exp.	plural	zoots	s	2	left	yellow	group	right	purple	ind.	ACW
8	1	exp.	plural	tulbs	z	1	left	green	group	right	red	ind.	CW
9	2	exp.	singular	feek		3	left	green	ind.	right	red	group	CW
10	2	exp.	singular	glaive		4	left	yellow	ind.	right	purple	group	ACW
11	2	exp.	singular	zoot		2	right	purple	ind.	left	yellow	group	ACW
12	2	exp.	singular	tulb		1	right	red	ind.	left	green	group	CW
13	2	exp.	plural	feeks	s	3	right	red	group	left	green	ind.	CW
14	2	exp.	plural	glaive	z	4	right	purple	group	left	yellow	ind.	ACW
15	2	exp.	plural	zoots	s	2	left	yellow	group	right	purple	ind.	ACW
16	2	exp.	plural	tulbs	z	1	left	green	group	right	red	ind.	CW
17	1	control	singular	chair		5	left	blue	ind.	right	green	group	CW
18	1	control	plural	spoon		6	right	green	group	left	blue	ind.	ACW
19	2	control	singular	spoon		6	left	blue	ind.	right	green	group	ACW
20	2	control	plural	chairs		5	left	green	group	right	red	ind.	CW

Experiment 2: List D

No	Cond	Trial type	Number	Label	Allo-morph	Obj.	Target Pos.	Target Colour	Target type	Comp. pos.	Comp. Colour	Comp Type	Turn Direction
1	1	exp.	singular	feek		3	right	green	ind.	left	red	group	CW
2	1	exp.	singular	glaive		4	right	yellow	ind.	left	purple	group	ACW
3	1	exp.	singular	zoot		2	left	purple	ind.	right	yellow	group	ACW
4	1	exp.	singular	tulb		1	left	red	ind.	right	green	group	CW
5	1	exp.	plural	feeks	s	3	left	red	group	right	green	ind.	CW
6	1	exp.	plural	glaive	z	4	left	purple	group	right	yellow	ind.	ACW
7	1	exp.	plural	zoots	s	2	right	yellow	group	left	purple	ind.	ACW
8	1	exp.	plural	tulbs	z	1	right	green	group	left	red	ind.	CW
9	2	exp.	singular	feek		3	right	green	ind.	left	red	group	CW
10	2	exp.	singular	glaive		4	right	yellow	ind.	left	purple	group	ACW
11	2	exp.	singular	zoot		2	left	purple	ind.	right	yellow	group	ACW
12	2	exp.	singular	tulb		1	left	red	ind.	right	green	group	CW
13	2	exp.	plural	feeks	s	3	left	red	group	right	green	ind.	CW
14	2	exp.	plural	glaive	z	4	left	purple	group	right	yellow	ind.	ACW
15	2	exp.	plural	zoots	s	2	right	yellow	group	left	purple	ind.	ACW
16	2	exp.	plural	tulbs	z	1	right	green	group	left	red	ind.	CW
17	1	control	singular	chair		5	right	blue	ind. dual	left	green	group	CW
18	1	control	plural	spoon		6	left	green	group	right	blue	ind.	ACW
19	2	control	singular	spoon		6	right	blue	ind. dual	left	green	group	ACW
20	2	control	plural	chairs		5	right	green	group	left	red	ind.	CW

Experiment 3: List A

Congruent trials

No	Cond	Trial type	Number	Label	Allo-morph	Obj	Target Pos.	Target Colour	Target Type	Comp. Pos.	Comp. Colour	Comp. Type	Turn Direction
1	1	exp.	singular	feek		1	right	yellow	group	left	purple	ind.	CW
2	1	exp.	singular	glaive		2	right	green	group	left	red	ind.	ACW
3	1	exp.	singular	zoot		3	left	red	group	right	green	ind.	ACW
4	1	exp.	singular	tulb		4	left	purple	group	right	yellow	ind.	CW
5	1	exp.	plural	feeks	s	1	left	purple	ind.	right	yellow	group	CW
6	1	exp.	plural	glaives	z	2	left	red	ind.	right	green	group	ACW
7	1	exp.	plural	zoots	s	3	right	green	ind.	left	red	group	ACW
8	1	exp.	plural	tulbs	z	4	right	yellow	ind.	left	purple	group	CW
17	1	control	singular	chair		5	right	green	group	left	blue	ind.	CW
18	1	control	plural	spoons	z	6	left	blue	ind.	right	green	group	ACW

Incongruent trials

No	Cond	Trial type	Number	Label	Allo-morph	Obj	Gesture target			Sentence target			Turn
							pos.	col.	type.	pos.	col.	type	
9	2	exp.	singular	feek		1	right	yellow	group	left	purple	ind.	CW
10	2	exp.	singular	glaive		2	right	green	group	left	purple	ind.	ACW
11	2	exp.	singular	zoot		3	left	red	group	right	yellow	group	ACW
12	2	exp.	singular	tulb		4	left	purple	group	right	green	group	CW
13	2	exp.	plural	feeks	s	1	left	purple	ind.	right	yellow	group	CW
14	2	exp.	plural	glaives	z	2	left	red	ind.	right	green	group	ACW
15	2	exp.	plural	zoots	s	3	right	green	ind.	left	purple	ind.	ACW
16	2	exp.	plural	tulbs	z	4	right	yellow	ind.	left	purple	ind.	CW
19	2	control	singular	chair		5	right	green	group	left	blue	group	CW
20	2	control	plural	spoons	z	6	left	blue	group	right	green	group	ACW

Experiment3: List B

Congruent trials

No	Cond	Trial type	Number	Label	Allo-morph	Obj	Target Pos.	Target Colour	Target Type	Comp. Pos.	Comp. Colour	Comp. Type	Turn Direction
1	1	exp.	singular	feek		1	left	yellow	group	right	purple	ind.	CW
2	1	exp.	singular	glaive		2	left	green	group	right	red	ind.	ACW
3	1	exp.	singular	zoot		3	right	red	group	left	green	ind.	ACW
4	1	exp.	singular	tulb		4	right	purple	group	left	yellow	ind.	CW
5	1	exp.	plural	feeks	s	1	right	purple	ind.	left	yellow	group	CW
6	1	exp.	plural	glaives	z	2	right	red	ind.	left	green	group	ACW
7	1	exp.	plural	zoots	s	3	left	green	ind.	right	red	group	ACW
8	1	exp.	plural	tulbs	z	4	left	yellow	ind.	right	purple	group	CW
17	1	control	singular	chair		5	left	green	group	right	blue	ind.	CW
18	1	control	plural	spoons	z	6	right	blue	ind.	left	green	group	ACW

Incongruent trials

No	Cond	Trial type	Number	Label	Allo-morph	Obj	Gesture target			Sentence target			Turn
							pos.	col.	type.	pos.	col.	type	
9	2	exp.	singular	feek		1	left	yellow	group	right	purple	ind.	CW
10	2	exp.	singular	glaive		2	left	green	group	right	purple	ind.	ACW
11	2	exp.	singular	zoot		3	right	red	group	left	yellow	group	ACW
12	2	exp.	singular	tulb		4	right	purple	group	left	green	group	CW
13	2	exp.	plural	feeks	s	1	right	purple	ind.	left	yellow	group	CW
14	2	exp.	plural	glaives	z	2	right	red	ind.	left	green	group	ACW
15	2	exp.	plural	zoots	s	3	left	green	ind.	right	purple	ind.	ACW
16	2	exp.	plural	tulbs	z	4	left	yellow	ind.	right	purple	ind.	CW
19	2	control	singular	chair		5	left	green	group	right	blue	group	CW
20	2	control	plural	spoons	z	6	right	blue	group	left	green	group	ACW

Experiment 3: List C

Congruent trials

No	Cond	Trial type	Number	Label	Allo-morph	Obj	Target Pos.	Target Colour	Target Type	Comp. Pos.	Comp. Colour	Comp. Type	Turn Direction
1	1	exp.	singular	feek		3	right	yellow	group	left	purple	ind.	CW
2	1	exp.	singular	glaive		4	right	green	group	left	red	ind.	ACW
3	1	exp.	singular	zoot		1	left	red	group	right	green	ind.	ACW
4	1	exp.	singular	tulb		2	left	purple	group	right	yellow	ind.	CW
5	1	exp.	plural	feeks	s	3	left	purple	ind.	right	yellow	group	CW
6	1	exp.	plural	glaives	z	4	left	red	ind.	right	green	group	ACW
7	1	exp.	plural	zoots	s	1	right	green	ind.	left	red	group	ACW
8	1	exp.	plural	tulbs	z	2	right	yellow	ind.	left	purple	group	CW
17	1	control	singular	chair		5	right	green	group	left	blue	ind.	CW
18	1	control	plural	spoons	z	6	left	blue	ind.	right	green	group	ACW

Incongruent trials

No	Cond	Trial type	Number	Label	Allo-morph	Obj	Gesture target			Sentence target			Turn
							pos.	col.	type.	pos.	col.	type	
9	2	exp.	singular	feek		3	right	yellow	group	left	purple	ind.	CW
10	2	exp.	singular	glaive		4	right	green	group	left	purple	ind.	ACW
11	2	exp.	singular	zoot		2	left	red	group	right	yellow	group	ACW
12	2	exp.	singular	tulb		2	left	purple	group	right	green	group	CW
13	2	exp.	plural	feeks	s	3	left	purple	ind.	right	yellow	group	CW
14	2	exp.	plural	glaives	z	4	left	red	ind.	right	green	group	ACW
15	2	exp.	plural	zoots	s	1	right	green	ind.	left	purple	ind.	ACW
16	2	exp.	plural	tulbs	z	2	right	yellow	ind.	left	purple	ind.	CW
19	2	control	singular	chair		5	right	green	group	left	blue	group	CW
20	2	control	plural	spoons	z	6	left	blue	group	right	green	group	ACW

Experiment3: List D

Congruent trials

No	Cond	Trial type	Number	Label	Allo-morph	Obj	Target Pos.	Target Colour	Target Type	Comp. Pos.	Comp. Colour	Comp. Type	Turn Direction
1	1	exp.	singular	feek		3	left	yellow	group	right	purple	ind.	CW
2	1	exp.	singular	glaive		4	left	green	group	right	red	ind.	ACW
3	1	exp.	singular	zoot		1	right	red	group	left	green	ind.	ACW
4	1	exp.	singular	tulb		2	right	purple	group	left	yellow	ind.	CW
5	1	exp.	plural	feeks	s	3	right	purple	ind.	left	yellow	group	CW
6	1	exp.	plural	glaives	z	4	right	red	ind.	left	green	group	ACW
7	1	exp.	plural	zoots	s	1	left	green	ind.	right	red	group	ACW
8	1	exp.	plural	tulbs	z	2	left	yellow	ind.	right	purple	group	CW
17	1	control	singular	chair		5	left	green	group	right	blue	ind.	CW
18	1	control	plural	spoons	z	6	right	blue	ind.	left	green	group	ACW

Incongruent trials

No	Cond	Trial type	Number	Label	Allo-morph	Obj	Gesture target			Sentence target			Turn
							pos.	col.	type.	pos.	col.	type	
9	2	exp.	singular	feek		3	left	yellow	group	right	purple	ind.	CW
10	2	exp.	singular	glaive		4	left	green	group	right	purple	ind.	ACW
11	2	exp.	singular	zoot		1	right	red	group	left	yellow	group	ACW
12	2	exp.	singular	tulb		2	right	purple	group	left	green	group	CW
13	2	exp.	plural	feeks	s	3	right	purple	ind.	left	yellow	group	CW
14	2	exp.	plural	glaives	z	4	right	red	ind.	left	green	group	ACW
15	2	exp.	plural	zoots	s	1	left	green	ind.	right	purple	ind.	ACW
16	2	exp.	plural	tulbs	z	2	left	yellow	ind.	right	purple	ind.	CW
19	2	control	singular	chair		5	left	green	group	right	blue	group	CW
20	2	control	plural	spoons	z	6	right	blue	group	left	green	group	ACW

Appendix 4: Additional analyses

1. Experiment 1

1.1 Reaction time

1.1.1 Analysis

If the amount of attention participants pay to the target increases post stimulus, they should be quicker to orient their attention to the target than to the competitor. That is, after exposure to the stimulus they should look at the target before they look at the competitor. This variable therefore records the time between the onset of the post-stimulus window and the first shift-to-group (group latency) and the time between the onset of the post-stimulus window and the first shift-to-individual (individual latency). The group latency is then subtracted from the individual latency to give a difference score. A positive difference score indicates that the first shifts tended to be shifts-to-group. A negative difference score indicates that the first shifts tended to be shifts-to-individual. As previously, the onset of the post-stimulus window was 300ms after the offset of the noun. The window lasted until the end of the trial.[48] Trials on which the participants were already looking at the group or the individual at the onset of the post-stimulus window and trials on which the participants did not shift their attention to both the group and the individual during the post-stimulus window were excluded from further analysis. Following this trial loss, 9 participants failed to contribute data to all four conditions. Their data was excluded. Combined with the trials lost from the remaining participants this left only 196 trials (49%) across 24 participants (73%) available for analysis. It was decided that this was not enough data to pursue this analysis further.

1.2 Shift Probability

1.2.1 Analysis

If the amount of attention participants pay to the group increases after observing the plural stimuli, they should be more likely to shift their attention to the group than to the individual during the post-stimulus window of plural trials. That is, there should be a higher probability of shifts-to-group than shifts-to-individual. Likewise, if the amount of attention participants pay to the individual increases after observing the singular stimuli, they should be more likely to shift their attention to the individual than to the group during the post-stimulus window of singular trials. That is, there should be a higher probability of shifts-to-individual than shifts-to-group. The dependent variable for this

[48] The longest window possible was used to include as many trials as possible in which the participant looks at both the group and the individual post-stimulus.

analysis is onset contingent, i.e. it is calculated separately for trials on which the participant is looking at the group at the onset of the post-stimulus window (group-initial trials) and those on which the participant is looking at the individual at this point (individual-initial trials). For group-initial trials, the dependent variable records looks-to-individual divided by looks-to-individual plus looks-to-group. This reflects the proportion of trials on which participants shifted their attention from the group to the individual. If participants are sensitive to the stimuli, the shift probability on group-initial trials should therefore be higher on singular trials than plural trials. Likewise, for individual-initial trials, the dependent variable records looks-to-group divided by looks-to-group plus looks-to-individual. This reflects the proportion of trials on which participants shifted their attention from the individual to the group. If participants are sensitive to the stimuli, shift probability on individual-initial trials should therefore be higher on plural trials than on singular trials. This analysis excludes all trials on which the participant is not looking at either the group or the individual at the onset of the post-stimulus window. It also excludes all trials in which the participant does not look at either the group or the individual in the post stimulus window (as otherwise the ratio cannot be calculated). In this analysis there are eight experimental conditions (see Table 24) Following the exclusions, 29 participants failed to contribute data to all conditions. Their data was excluded. Combined with the trial loss from the remaining participants this left only 68 trials (17%) across 4 participants (12%). I decided this was not enough data to pursue this analysis further.

Table 24: List of conditions in onset contingent shift probability analysis, Experiment 1

Condition No.	Description
1	Group-initial singular trials with nominal number marking
2	Group-initial singular trials with nominal and verbal number marking
3	Group-initial plural trials with nominal number marking
4	Group-initial plural trials with nominal and verbal number marking
5	Individual-initial singular trials with nominal number marking
6	Individual-initial singular trials with nominal and verbal number marking
7	Individual-initial plural trials with nominal number marking
8	Individual-initial plural trials with nominal and verbal number marking

2. Experiment 2

2.1 Reaction time

2.1.1 Analysis

If the amount of attention participants pay to the target increases after observing the stimuli, they should look at the target before the competitor in the post-stimulus window. The dependent variable for this analysis therefore records the time between the onset of the post-stimulus window and the first shift-to-target (target latency) and the time between the onset of the post-stimulus window and the first shift-to-competitor (competitor latency). The target latency is then subtracted from the competitor latency to give a difference score. A positive difference score indicates that the first shifts tended to be shifts-to-target. A negative difference score indicates that the first shifts tended to be shifts-to-competitor. As before, the post-stimulus window began 300ms after the onset of the stimulus (i.e. at 7300ms). The window continued until the end of the trial.[49] Trials on which the participants were already looking at the target or the competitor at the onset of the post-stimulus window and trials on which the participants do not look at both the target and the competitor post-stimulus were excluded from this analysis.[50] If participants are sensitive to both gestures the difference score should be significantly higher than zero for both Condition 1 and Condition 2. Furthermore, if participants are more sensitive to body orientation and pointing than body orientation alone, the difference score should be higher in Condition 2 than Condition 1.

2.1.2 Results

The descriptive statistics are given in Table 25. All values are within 2.5 standard deviations of the mean.

[49] As before, the longest window possible was used to include as many trials as possible in which the participant looks at both the target and the competitor post-stimulus.
[50] One hundred and ninety-one trials were excluded in this way. This led to a loss of power, but allowed for the calculation of the dependent variable. Following this trial loss, 1 participant failed to contribute data to both conditions and their data were excluded from this analysis. This analysis was therefore run with the data from remaining 40 participants, totalling 312 trials (72%).

Table 25: Difference score means (ms) by condition and age, reaction time analysis, Experiment 1

	Mean (SD)	**Mean (SD) by age**	
		2yos	3yos
Cond.1	-189 (1456)	-42.1 (1283)	-351 (1646)
Cond.2	624 (1159)	560 (1013)	694 (1327)

As in the previous analyses, I used lme4 package (Bates et al., 2012) for R (R Core Team, 2019) to perform a linear mixed effects analysis. The model was constructed in the same way as before; the fixed effects were condition and age (added with interaction terms) and the random effect was participant. Again, the model included a by-participant random slope for the effect of condition.[51] A summary of the results is given in Table 26.

Table 26: Results of a linear mixed effects model examining the effect of condition and age on difference score, reaction time analysis, Experiment 1

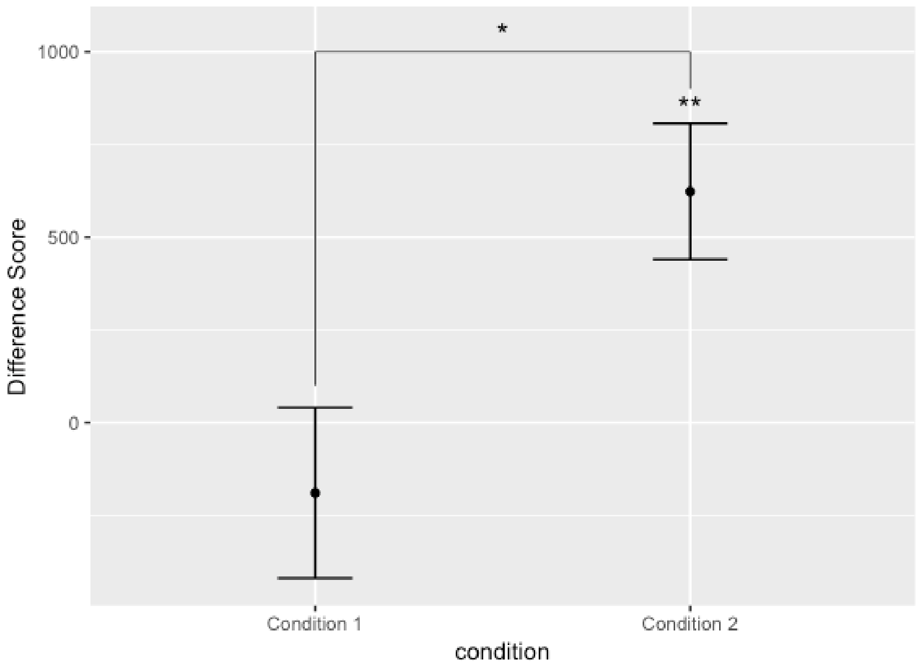
Factor	**Estimate**	**Std. Error**	**df**	***t***	***p***	
Condition	599.53	212.40	33.95	2.823	<.01	**
Age	14.27	16.48	35.25	0.866	0.39	
Condition:Age	14.24	32.35	31.63	0.440	0.66	

As shown, the results show a significant effect of condition, no effect of age and no significant interaction between the two. To confirm that age does not significantly affect difference score, I created a model that did not include age as a fixed effect.[52] I then used ANOVA to compare the full and reduced models. The results showed that the full model was no better a fit than the reduced model (log likelihood: -3556 versus -3557, ($\chi 2(2) = 1.07$, $p = .59$). As such, I proceeded with the full model. To determine whether condition has a significant impact on difference score or not, I created a model with condition and its associated slope removed[53] and compared this model to the model with this variable included. The results showed the model that included condition was a significantly better fit than the reduced model (log likelihood: -3557 versus -3561, ($\chi 2(2) = 9.17$, $p < .05$). This suggests condition has a significant impact on difference score.

[51] Formula: model.full = lmer(diffscore ~ condition * age + (1 + condition | participant), data=exp1.rt)
[52] Formula: lmer(diffscore ~ condition + (1 + condition | participant), data=exp1.rt)
[53] Formula: lmer(diffscore ~ age + (1| participant), data=exp1.rt)

In particular, the difference score in Condition 2 (*M* = 624) is significantly higher than the difference score in Condition 1(*M* = -189). These differences are illustrated in Figure 22. These results show that after observing the gestures, participants tended to look at the target before they looked at the competitor in Condition 2, but not Condition 1. The results of the reaction time analysis therefore support the results of the proportional looks and longest looks analyses in showing that participants attend to body orientation and pointing but not to body orientation alone.

Figure 22: Difference score (ms) by condition, reaction time analysis, Experiment 1



Reaction time analyses typically use target latency (rather than difference score) as their dependent measure. To determine whether this more traditional approach yields different results, I ran a follow-up analysis using target latency as the dependent measure. Target latency was defined as the time between the onset of the post-stimulus window and the first look-to-target. The results of this second analysis support those of the first. Specifically, model comparison showed no effect of age and a significant effect of condition such that the target latency was significantly shorter in Condition 2 (*M* = 1925ms) than Condition 1 (*M* = 2501ms, $t(39) = 3.54$, $p < .01$). This suggests that participants looked at the target more quickly in Condition 2 than in Condition 1. Again, this supports previous results showing that the body orientation and pointing gesture was more effective in directing attention to the target than the body orientation only gesture.

2.2 Shift Probability

2.2.1 Analysis

If the amount of attention participants pay to the target increases after observing the stimuli, they should be more likely to shift their attention to the target than to the competitor during the post-stimulus window. That is, there should be a higher probability of shifts-to-target than shifts-to-competitor. The dependent variable for a shift probability analysis is onset contingent, i.e. it is calculated separately for trials on which the participant is looking at the target at the onset of the post-stimulus window and those on which the participant is looking at the competitor at this point. For target-initial trials, the dependent variable is looks-to-competitor divided by looks-to-competitor plus looks-to-target. This reflects the proportion of trials on which participants shifted their attention from the target to the competitor. For competitor-initial trials, the dependent variable is looks-to-target divided by looks-to-target plus looks-to-competitor. Again, this reflects the proportion of trials on which participants shifted their attention, this time from the competitor to the target. This analysis excludes all trials on which the participants were not looking at either the target or the competitor at the onset of the post-stimulus window. It also excludes all trials in which the participants do not look at either the target or the competitor in the post stimulus window (as otherwise the ratio cannot be calculated). Following these exclusions, 23 participants failed to contribute data to all four conditions (see Table 27). Their data was excluded. Combined with the trials lost from the remaining participants this left only 151 trials (35%) across 18 participants (44%) available for analysis. I decided this was not enough data to pursue this analysis further.

Table 27: List of conditions in onset contingent shift probability analysis, Experiment 2

Condition No.	**Description**
1	Target-initial trials with the body orientation cue
2	Target-initial trials with the body orientation and pointing cue
3	Competitor-initial trials with the body orientation cue
4	Competitor-initial trials with the body orientation and pointing cue

3. Experiment 3

3.1 Reaction Time

If the amount of attention participants pay to the group increases after observing the stimuli, they should look at the group before the individual in the post-stimulus window. The dependent variable in this analysis records the time between the onset of the post-stimulus window and the first shift-to-group (group latency) and the time between the onset of the post-stimulus window and the first shift-to-individual (individual latency). The group latency is then subtracted from the individual latency to give a difference score. A positive difference score indicates that the first shifts tended to be shifts-to-group. A negative difference score indicates that the first shifts tended to be shifts-to-individual. Trials on which participants were already looking at the group or the individual at the onset of the post-stimulus window and trials on which the participants did not shift their attention to both the group and the individual were excluded from further analysis. Following this trial loss, 16 participants failed to contribute contributed data to all four conditions. Combined with the excluded data from the remaining participants this left only 142 trials (36%) across 18 participants (53%) available for analysis. I decided not to pursue this analysis further.

3.2 Shift Probability

If the amount of attention participants pay to the group increases after observing the plural stimuli, they should be more likely to shift their attention to the group than to the individual during the post-stimulus window of plural trials. That is, there should be a higher probability of shifts-to-group than shifts-to-individual. Likewise, if the amount of attention participants pay to the individual increases after observing the singular stimuli, they should be more likely to shift their attention to the individual than to the group during the post-stimulus window of singular trials. That is, there should be a higher probability of shifts-to-individual than shifts-to-group. The dependent variable for this analysis is onset contingent, i.e. it is calculated separately for trials on which the participant is looking at the group at the onset of the post-stimulus window (group-initial trials) and those on which the participant is looking at the individual at this point (individual-initial trials). On group-initial trials shifts in attention indicate shifts to the individual. If participants are sensitive to the stimuli, the shift probability on group-initial trials should be higher on singular trials than plural trials. Likewise, on individual-initial trials shifts in attention indicate shifts to the group. If participants are sensitive to the stimuli shift probability on individual-initial trials should be higher on plural trials than on singular trials. This analysis excludes all trials on which the participant is not looking at either the group or the individual at the onset of the post-stimulus window. It also excludes trials on which the participants looked at neither the group nor the individual in the post-stimulus window. Following these exclusions, no participants contributed data to all eight conditions (see Table 28). This analysis was therefore abandoned.

Table 28: List of conditions in onset contingent shift probability analysis, Experiment 3

Condition No.	**Description**
1	Group-initial singular trials in the congruent condition
2	Group-initial singular trials in the incongruent condition
3	Group-initial plural trials in the congruent condition
4	Group-initial plural trials in the incongruent condition
5	Individual-initial singular trials in the congruent condition
6	Individual-initial singular trials in the incongruent condition
7	Individual-initial plural trials in the congruent condition
8	Individual-initial plural trials in the incongruent condition

References

Akhtar, N. (2005). The robustness of learning through overhearing. *Developmental Science, 8*(2), 199-209.
Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development, 67*(2), 635-645. doi:10.1111/1467-8624.ep9605280332
Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child Development, 72*(2), 416-430.
Akhtar, N., & Tomasello, M. (2001). The social nature of words and word-learning. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, N. Akhtar, M. Tomasello, & G. J. Hollich (Eds.), *Becoming a word learner: a debate on lexical acquisition*. New York, NY: Oxford University Press.
Bahrick, L. E., Lickliter, R., & Flom, R. A. (2004). Intersensory redundancy guides the development of selective attetnion, perception and cognition in infancy. *Current Directions in Psychological Science, 13*(3), 99-102.
Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development, 62*(5), 875-890.
Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language, 20*(02), 395-418.
Baldwin, D. A., & Moses, L. J. (2001). Links between social understanding and early word learning: challenges to current accounts. *Social Development, 10*(3), 309-329.
Bates, D. M., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0.
Berko, J. (1958). The child's learning of English morphology *Word, 14*, 150-177.
Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.
Bloom, P., & Keleman, D. (1996). Syntactic cues in the acquisition of collective nouns. *Cognition, 56*(1), 1-30.
Blossom, M. (2013). *Assessing two-year-old's knowledge of number agreement morphology* (Doctor of Philosophy), University of Kansas.
Briganti, A. M., & Cohen, L. B. (2011). Examining the role of social cues in early word learning. *Infant Behavior and Development, 34*(1), 211-214.
Brooks, R., & Meltzoff, A. N. (2002). The importance of eyes: how infants interpret adult looking behavior. *Developmental Psychology, 38*(6), 958-966.
Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science, 8*(6), 535-543.
Brown, R. W. (1957). Linguistic determinism and the part of speech. *Journal of Abnormal and Social Psychology, 55*, 1-5.
Brown, R. W. (1973). *A first language: The early stages*. Cambridge, MA: Havard Univeristy Press.
Butterworth, G. E., & Grover. (1988). The origins of referential communication in human infancy. In L. Weiskrantz (Ed.), *Thought without language* (pp. 5-24). Oxford, UK: Oxford University Press.
Butterworth, G. E., & Grover. (1990). Joint visual attention, manual pointing, and preverbal communication in human infancy. In M. Jeannerod (Ed.), *Attention and performance XIII: motor representation and control* (pp. 605-624). Hillsdale, NJ: Lawrence Erlbaum.
Butterworth, G. E., & Jarrett, N. (1991). What minds have in common is space: spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology, 9*(1), 55-72.

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen- through 18-month old infants differentially imitate intentional and accidental actions *Infant Behavior and Development, 21*(2), 315-330.
Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(4), 1-143. doi:10.2307/1166214
Chow, V., Poulin ‐ Dubois, D., & Lewis, J. (2008). To see or not to see: Infants prefer to follow the gaze of a reliable looker. *Developmental Science, 11*(5), 761-770.
Christophe, A., Milotte, S., Bernal, S., & Lidz, J. (2008). Bootstrapping lexical and syntactic acquisition. *Language & Speech,, 51*(1), 61-75.
Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2004). *Habit X: A new program for obtaining and organizing data in infant perception and cognition studies*. Austin: University of Texas.
Corkum, V., & Moore, C. (1995). Development of joint visual Attention in infants. In C. Moore & P. J. Dunham (Eds.), *Joint attention: its origins and role in development*. Hillsdale, NJ: Lawrence Erlbaum Associates.
D'Entremont, B., Hains, S. M. J., & Muir, D. W. (1997). A demonstration of gaze following in 3- to 6-months-old *Infant Behavior and Development, 20*(4), 569-572.
Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers, 28*(1), 125-127.
Davies, B., Xu Rattanasone, N., & Demuth, K. (2016). Two-year-olds’ sensitivity to inflectional plural morphology: allomorphic effects. *Language Learning and Development*, 1-16.
Davies, B., Xu Rattanasone, N., & Demuth, K. (2019). Acquiring the last plural: Morphological effects on the acquisition of /-əz/. *Manuscript submitted for publication*.
Davies, B., Xu Rattanasone, N., Schembri, T., & Demuth, K. (2019). Preschoolers' developing comprehension of the plural: The effects of number and allomorphic variation *Journal of Experimental Child Psychology, 185*, 95-108.
de León, L. (2011). Language socialization and multiparty participation frameworks. In A. Duranti, E. Ochs, & B. B. Schieffelin (Eds.), *The handbook and language socialisation*. Malden: Wiley-Blackwell.
Deák, G. O., Flom, R. A., & Pick, A. D. (2000). Effects of gesture and target on 12- and 18-month-olds' joint visual attention to objects in front of or behind them. *Developmental Psychology, 36*(4), 511-523.
Dunphy-Lelii, & Wellman. (2004). Infants' understanding of occlusion of others' line-of-sight: Implications for an emerging theory of mind. *European Journal of Developmental Psychology, 1*(1), 49-66.
Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye-movements to monitor spoken language comprehension by infants and young children. In I. A. Sekerina, E. M. Fernández, & H. Clahsen (Eds.), *Developmental psycholinguistics : on-line methods in children's language processing* (pp. 184-218). Amsterdam; Philadelphia: John Benjamins Pub.
Fisher, C. (2002). Structural limits on verb mapping: the tolel of abstract structure in 2.5-year-olds' interpretations of novel verbs. *Developmental Science, 5*(1), 55-64.
Gleitman, L. R. (1990). The Structural Sources of Verb Meanings. *Language Acquisition, 1*(1), 3-55.
Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard Words. *Language Learning and Development, 1*(1), 26-64.
Gleitman, L. R., & Landau, B. (1994). *The acquisition of the lexicon*. Cambridge, Mass: CogNet.
Gogate, L., & Bahrick, L. E. (1998). Intersensory redundancy facilitates learning of arbitary relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology, 69*, 133-149.

Golinkoff, & Hirsh-Pasek. (2000). *How babies talk: the magic and mystery of language in the first three years of life*. New York, NY: Plume.

Golinkoff, & Hirsh-Pasek. (2006). Baby Wordsmith: From Associationist to Social Sophisticate. *Current Directions in Psychological Science, 15*(1), 30-33.

Golinkoff, Hirsh-Pasek, Cauley, & Gordon. (1987). The eyes have it: lexical and syntactic comprehension in a new paradigm. *Journal of Child Language, 14*, 23-55.

Gordon, P. (2003). *The origin of argument stucture in infant event representations. .* Paper presented at the 28th Annual Boston University Conference on Language Development, Boston, MA.

Grassmann, S. (2014). The pragmatics of word learning In D. Matthews (Ed.), *Pragmatic development in first language acquisition*. Philadelphia, PA: John Benjamns Publishing Company.

Grassmann, S., Stracke, M., & Tomasello, M. (2009). Two-year-olds exclude novel objects as potential referents of novel words based on pragmatics. *Cognition, 112*(3), 488-493.

Grassmann, S., & Tomasello, M. (2010). Young children follow pointing over words in interpreting acts of reference. *Developmental Science, 13*(1), 252-263.

Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology, 53*(4), 310-344.

Hall, D. G., Lee, S. C., & Belanger, J. (2001). Young Children's Use of Syntactic Cues to Learn Proper Names and Count Nouns. *Developmental Psychology, 37*(3), 298-307.

Hall, D. G., Waxman, S. R., & Hurwitz, W. M. (1993). How two- and four-year-old children interpret adjectives and count nouns. *Society for Research in Child Development,, 64*(6), 1651-1664.

Hirsh-Pasek, K., & Golinkoff, R. M. (2001). The Whole is Greater than the Sum of the Parts or Why the Emergentist Coalition Model Works. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello, & G. J. Hollich (Eds.), *Becoming a Word Learner: A Debate on Lexical Aqcuisition*. New York, NY: Oxford University Press

Hirsh-Pasek, K., & Golinkoff, R. M. (2008). King Solomon's Take on Word Learning: An Integrative Account From the Radical Middle. *Advances in Child Development and Behaviour, 15*, 30-33.

Hirsh-Pasek, K., Golinkoff, R. M., & Hollich, G. J. (2001). An Emergentist Coalition Model for Word-Learning: Mapping Words to Objects is Product of Multiple Cues. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello, & G. J. Hollich (Eds.), *Becoming a Word Learner: A Debate on Lexical Acquisition*. New York, NY: Oxford University Press.

Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., . . . Bloom, L. (2000). Breaking the language barrier: an emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development, 65*(3), i-135.

Houston-Price, C., Plunkett, K., & Duffy, H. (2006). The use of social and salience cues in early word learning. *Journal of Experimental Child Psychology, 95*(1), 27-55.

Huttenlocher, J., Smiley, P., & Charney, R. (1983). Emergence of action categories in the child: evidence from verb meanings. *Psychological Review, 90*, 72-93.

Jaswal, V., & Hansen, M. B. (2006). Learning words: children disregard some pragmatic information that conflicts with mutual exclusivity. *Developmental Science, 9*(2), 158-165.

Jaswal, V., & Markman, E. (2001). Learning Proper and Common Names in Inferential versus Ostensive Contexts. *Child Development, 72*(3), 768.

Jolly, H. R., & Plunkett, K. (2008). Inflectional bootstrapping in 2-year-olds. *Language & Speech, 51*(1/2), 45-59.
Katz, N., Baker, E., & Macnamara, J. (1974). What's in a Name? A Study of How Children Learn Common and Proper Names. *Child Development, 45*(2), 469-473.
Kouider, S., Halberda, J., Wood, J., & Carey, S. (2006). Aqcuisition of English number marking: the singular-plural distinction *Language Learning and Development, 2*(1), 1-25.
Küntay, A., Nakamura, K., & Ateş Şen, B. (2014). Crosslinguistic and crosscultrual approaches to pragmatic development. In D. Matthews (Ed.), *Pragmatic development in first language acquisition* Philadelphia, PA: John Benjamins Publishing Company.
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-177.
Lempers, J., D. (1979). Young children's production and comprehension on nonverbal deictic behaviours. *Journal of Genetic Psychology, 135*(1), 93-102.
Lieven, E., & Stoll, S. (2013). Early communicative development in two cultures: a comparison of the communicative environments of children from two cultures *Human Development, 56*(3), 178-206.
Lukyanenko, C. (2011). *Where are the cookies? Three-year-olds use number-marked verbs to anticipate upcoming nouns.* (Master of Arts), University of Illinois at Urbana-Champaign.
Lukyanenko, C., & Fisher, C. (2016). Where are the cookies? Two- and three-year-olds use number-marked verbs to anticipate upcoming nouns. *Cognition, 146*, 349-370.
Mastin, J. D., & Vogt, P. (2016). Infant engagement and early vocabulary development: a naturalistic observation study of Mozambican infants from 1;1 to 2;1. *Journal of Child Language, 43*(2), 235-264.
Merriman, W. E., & Bowman, L. L. (1989). The mutual exclusivity bias in children's word learning. *with discussion, 54*(3-4), 1-129.
Moll, H., & Tomasello, M. (2004). 12 ‐ and 18 ‐ month ‐ old infants follow gaze to spaces behind barriers. *Developmental Science, 7*(1), F1-F9.
Moore, C. (2008). The development of gaze following. *Child Development Perspectives, 2*(2), 66-70.
Moore, C., Angelopoulos, M., & Bennet, P. (1999). Word learning in the context of referential and salience cues. *Developmental Psychology, 35*(1), 60-68.
Morales, M., Mundy, P., & Rojas, J., 373–377. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behavior and Development,, 21*, 373-377.
Naigles, L. R. (1990). Children use syntax to learn verb meanings. *Journal of Child Language, 17*(02), 357-374. doi:doi:10.1017/S0305000900013817
Naigles, L. R., & Kako, E. T. (1993). First Contact in Verb Acquisition: Defining a Role for Syntax. *Child Development, 64*(6), 1665-1687.
Nappa, R., Wessel, A., McEldoon, K. L., Gleitman, L. R., & Trueswell, J. C. (2009). Use of speaker gaze and syntax in verb learning. *Language Learning and Development, 5*, 203-234.
Ng, M. M., Demir, Ö. E., & So, W. C. (2015). The role of gesture in referential communication: a developmental perspective. In L. Serratrice & S. Allen (Eds.), *The Acquisition of Reference* Philadelphia, PA: John Benjamins Publishing Company.
Nurmsoo, E., & Bloom, P. (2008). Preschoolers' perspective taking in word learning: Do they blindly follow eye gaze? *Psychological Science, 19*(3), 211-215. doi:10.1111/j.1467-9280.2008.02069.x
Papafragou, A., Cassidy, K., & Gleitman, L. R. (2007). When we think about thinking: The acqusition of belief verbs. *Cognition, 105*(1), 125-165.

Paquette-Smith, M., & Johnson, E. K. (2016). Toddlers' use of grammatical and social cues to learn novel words. *Language Learning and Development, 12*(3), 328-337.
Piotroski, J., & Naigles, L. (2012). Intermodal preferential looking. In E. Hoff (Ed.), *Research methods in child language: a practical guide*. Oxford, UK: Blackwell Publishing Ltd.
Pruden, S. M., Hirsh-Pasek, K., Golinkoff, R. M., & Hennon, E. A. (2006). The birth of words: Ten-month-olds learn words through perceptual salience. *Child Development, 77*(2), 266-280. doi:10.1111/j.1467-8624.2006.00869.x
Quine, W. V. (1969). *Word and object*. Massachusetts USA: The M.I.T Press.
R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
Rader, N. d. V., & Zukow-Goldring, P. (2012). Caregivers’ gestures direct infant attention during early word learning: the importance of dynamic synchrony. *Language Sciences, 34*(5), 559-568.
Rowland, C. (2014). *Understanding child language acquisition*. Milton Park, Abingdon, Oxon: Routledge.
Samuelson, L. K., & Smith, L. B. (1998). Memory and attention make smart word learning: An alternative account of Akhtar, Carpenter and Tomasello. *Child Development, 69*(1), 94-104.
Schafer, G., & Plunkett, K. (1998). Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child Development, 69*(2), 309-320.
Smith, L. B., Jones, S. S., & Landau, B. (1993). Counts nouns, adjectives and perceptual propoerties in children's novel word intepretations. *Developmental Psychology, 28*(2), 273-286.
Soja, N. N. (1992). Infernences about the meanings of nouns: The relationhsip between perception and syntax. *Cognitive Development, 7*(1), 29-45.
Sorrentino, C. M. (2001). Children and adults represent proper names as referring to unique individuals. *Developmental Science, 4*(4), 399-407.
Spelke, E. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology, 15*(626-636).
Stephens, G., & Matthews, D. (2014). The communicative infant from 0-18 months; the social-cognitive foundations of pragmatic development. In D. Matthews (Ed.), *Pragmatic development in first language acquisition* Philadelphia, PA: John Benjamins Publishing Company.
Swingley, D. (2012). The looking-while-listening procedure. In E. Hoff (Ed.), *Research methods in child language: a practical guide*. Oxford, UK: Blackwell Publishing Ltd.
Taylor, M., & Gelman, S. A. (1988). Adjective and nouns: Children's strategies for learning new words. *Child Development, 59*(2), 411-419.
Thomas, D. C., Campos, J., J., Shucard, W., Ransay, D. S., & Shucard, J. (1981). Semantic comprehension in infancy: a signal detecetion analysis. *Child Development, 52*, 798-803.
Tomasello, M. (1999). *The cultural origins of human cognition* Cambridge, MA: Harvard University Press.
Tomasello, M. (2001). Could we please lose the mapping metaphor, please? *Behavioral and Brain Sciences, 24*(06), 1119-1120.
Tomasello, M. (2003). *Constructing a language: a usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
Tomasello, M., & Barton, E. (1994). Learning words in non-ostenstive contexts. *Developmental Psychology, 30*(5), 639-650.
Tomasello, M., Hare, B., Lehmann, H., & Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis. *Journal of Human Evolution, 52*(3), 314-320.

Tomasello, M., Strosberg, R., & Akhtar, N. (1996). Eighteen-month-old children learn words in non-ostensive contexts. *Journal of Child Language, 23*, 157-176.
Werker, J. F., Cohen, L. B., Lloyd, V. L., & Casasola, M. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology, 34*(6), 1289-1309.
Wood, J. N., Kouider, S., & Carey, S. (2009). Acquisition of singular-plural morphology. *Developmental Psychology, 45*(1), 202-206.