Journal of Child Language (2019), **46**, 433–458
doi:10.1017/S0305000918000557

CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Learning speech-internal cues to pronoun interpretation from co-speech gesture: a training study

Whitney GOODRICH SMITH, Alexis K. BLACK, and Carla L. HUDSON KAM*

University of British Columbia, Canada
*Corresponding author: 2613 West Mall, Vancouver, British Columbia, Canada, V6T 1Z4.
E-mail: Carla.HudsonKam@ubc.ca

(Received 10 March 2017; revised 8 May 2018; accepted 15 November 2018;
first published online 18 January 2019)

Abstract

This study explores whether children can learn a structural processing bias relevant to pronoun interpretation from brief training. Over three days, 42 five-year-olds were exposed to narratives exhibiting a first-mentioned tendency. Two characters were introduced, and the first-mentioned was later described engaging in a solo activity. In our primary condition of interest, the Gesture Training condition, the solo-activity sentence contained an ambiguous pronoun, but co-speech gesture clarified the referent. There were two comparison conditions. In the Gender Training condition the characters were different genders, thereby avoiding ambiguity. In the Name Training condition, the first-mentioned name was simply repeated. Ambiguous pronoun interpretation was tested pre- and post-training. Children in the Gesture condition were significantly more likely to interpret ambiguous pronouns as the first-mentioned character after training. Results from the comparison conditions were ambiguous: there was a small but non-significant effect of training, but also no significant differences between conditions.

Keywords: pronoun interpretation; processing biases; training; gesture

Introduction

Research with adults shows the influence of a variety of things on the interpretation of third person pronouns. For example, adults have a tendency to interpret pronouns as referring to the entity introduced as the grammatical subject / first-mentioned character in previous discourse. This is an example of a bias that is related to the structure of the linguistic form, but listeners can use information outside of the linguistic signal as well; things like speaker's eye-gaze or co-speech gesture that can provide clues to the speaker's intended meaning also influence interpretation. Studies with children show that their sensitivity to cues to pronoun interpretation develops, suggesting that children take time to learn to use the cues to pronoun interpretation that adults are sensitive to.

© Cambridge University Press 2019

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

This study examines whether children can learn a structural bias within a short training period. Over several days of training we exposed children to narratives that exhibited a particular structural bias (the first-mentioned bias, described in detail below). The question was whether the children would extract the underlying pattern – the repeated reference to the first-mentioned character – during training, and the conditions under which they would do so. In our first condition, which was our primary condition of interest, the training narratives included ambiguous pronouns that were disambiguated by co-speech gesture produced by the speaker. We also had two comparison conditions. In the first comparison condition the training narratives also contained pronouns, but they were not ambiguous (the characters had different genders). In the second comparison condition there were no pronouns. Instead, the characters' names were used throughout. In all three conditions the first-mentioned character was referred to as the sole participant in a final activity, so all three exhibited the first-mentioned pattern.

The question was whether the children would extract the underlying pattern – the repeated reference to the first-mentioned character – and apply it to third person pronouns more broadly. The children's interpretations of ambiguous third person pronouns with no additional cue to interpretation were tested before and after training, and interpretation at the two time-points was compared to assess the effects of training. We found some evidence for pattern extraction in all three conditions, suggesting that concentrated exposure to narratives consistent with structural cues can help children learn to use such cues, even when the training narratives did not themselves contain the specific ambiguity being probed at test. However, the number of children who showed more bias after training than before was higher in the condition that contained the most relevant information, suggesting that training with ambiguous pronouns and non-structural cues to interpretation may be the most effective.

Pronoun interpretation in adults and children

Third person pronouns are technically ambiguous when they occur; they can refer to any entity that matches the pronoun in gender and number.[1] Despite this, adults generally fail to notice this ambiguity, and have little difficulty interpreting pronouns. Researchers have identified a variety of cues that influence pronoun interpretation in mature language users. For instance, adults have a preference for interpreting pronouns like 'he' and 'she' as referring to an antecedent previously introduced as a grammatical subject / first-mentioned character (Arnold, 2001; Järvikivi, van Gompel, Hyönä, & Bertram, 2005). That is, when more than one potential referent (i.e., referents matching the pronoun in number and gender) has been introduced into the discourse prior to the pronoun, adults are more likely to interpret the pronoun as being co-referential with the thing introduced as the grammatical subject / first. For instance, in (1), people tend to interpret *he* as being co-referential with Donald, not Mickey.

(1) "Donald is bringing some mail to Mickey, while a violent storm is beginning. He's carrying an umbrella, and it looks like they're both going to need it." (Arnold, Brown-Schmidt, & Trueswell, 2007, p. 549).

[1]The semantics of pronouns differ by language; not all languages' pronouns carry gender or number information and some carry other semantic features (e.g., animacy). This discussion focuses on English.

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

In English, subject-hood and being the first-mentioned entity tend to travel together – the grammatical subject is generally the first noun in the sentence – and so it can be difficult to know which property is important for interpretation. Various strands of evidence suggest that subjects are especially privileged when it comes to grammatical operations (see, e.g., Keenan, 1976), so it is reasonable to assume that subject-hood is the relevant factor for interpretation. Indeed, there is evidence to suggest that subjects are preferred as antecedents. Järvikivi *et al.* (2005) showed this for Finnish using an eye-tracking task. Finnish has free word order and uses case to indicate grammatical role. Järvikivi *et al.* had participants listen to sentences in SVO and OVS word order, followed by a sentence containing an ambiguous pronoun. In this way, they were able to disentangle grammatical role from order of mention. They found that participants were more likely to look at the character introduced as the grammatical subject upon hearing the pronoun regardless of the word order of the sentence in which the characters were introduced, that is, independent of whether the subject had been the first- or second-mentioned character.

But there is also evidence that being the first-mentioned noun contributes to this interpretation preference. Some of this evidence comes from the same study by Järvikivi *et al.* (2005). Although initial eye-movements were to the subject character, overall, the first-mentioned character attracted more looks than the second-mentioned character. A pure first-mentioned bias has been shown for English speakers as well. Goodrich Smith and Hudson Kam (2012) exposed adults to narratives containing two potential antecedents for an ambiguous pronoun that occurred later in the narrative. The potential antecedents were introduced in the same grammatical role as each other, with the result that subject-hood did not distinguish them, as the example stimulus item in (2) shows. After listening to the narration, participants were asked who had performed the action described in the pronoun-containing sentence. Responses showed that listeners were more likely to interpret the pronoun as referring to the first-mentioned than the second-mentioned character.

(2) "Annie and Sarah are having a picnic in the park. They have a lot of food with them. Annie is carrying the picnic basket, and Sarah has a blanket to sit on. She's excited about the cookies." (Goodrich Smith & Hudson Kam, 2012, p. 83).

The influence of subject-hood and order-of-mention on pronoun interpretation are only tendencies, and other things can influence pronoun interpretation such that the grammatical subject / first referent is not the preferred antecedent. For instance, when the pronoun is an accusative pronoun (e.g., *him* or *her*), grammatical objects are preferred as antecedents (Chambers & Smyth, 1998; Sauermann & Gagarina, 2017).

The previously mentioned effects are all structural, that is, they involve the structure of sentences in one way or another. However, other things can also affect interpretation preferences, for example, things like the semantics of the relevant nouns and verbs (see discussions in Pyykkönen & Järvikivi, 2010, and Pyykkönen, Matthews, & Järvikivi, 2010). The suggestion is that entities which are themselves more animate or whose animacy is made more salient (via their semantic role in the event described by the verb) are more cognitively accessible, and so more likely to be interpreted as the referent of an ambiguous pronoun.[2]

[2]Much of the adult literature is focused on understanding the mechanisms involved in pronoun interpretation, for instance whether it is about linguistic salience or general salience and so cognitive accessibility. See Arnold (2013) and Kehler and Rohde (2013) for discussions on these and other points.

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

As with the structural cues, semantic effects are in some sense built into the sentence. But there are other cues that are provided by the speaker themselves as to what their intended referent is. Stress is one such cue known to affect interpretation (Akmajian & Jackendoff, 1970). In (3), for example, whether *him* is interpreted as referring to Jacob or Bill shifts depending on stress. If the stress in the second clause is either neutral or on *Julie*, *him* is likely to be interpreted as Bill. However, if the stress is on *him*, people are more likely to interpret *him* as Jacob.

(3) Jacob bumped into Bill, then Julie bumped into him.

More recent work shows that cues to the speaker's intentions outside of the linguistic signal can also affect pronoun interpretations. For instance, eye-gaze and pointing directed at the intended referent can shift interpretation away from otherwise preferred referents (towards the thing looked at and pointed to; Nappa & Arnold, 2014). Co-speech gesture can also affect interpretation. When describing multiple entities people often gesture in different locations in space, establishing a metaphoric mapping between a specific character (or object) and a spatial location. This space is then available (for co-reference) when the speaker refers back to the same entity later in the discourse (Kendon, 2004; McNeill, 1992; So, Kita, & Goldin-Meadow, 2009). For instance, when talking about 'Sam' a speaker might place their palm up at their right side and then later repeat a similar gesture when mentioning Sam again. Goodrich Smith and Hudson Kam (2012) showed that listeners' pronoun interpretations are sensitive to these gestures. They showed participants narrations in which the speaker gestured to her left when mentioning one character and to her right when mentioning another. Later, the narrator produced a gesture alongside a pronoun that either went to the side previously associated with the first-mentioned character or the side previously associated with the second-mentioned character. When participants were asked who performed the action described in the sentence with the pronoun, participants were more likely to say that it was the second-mentioned character when they had seen a gesture consistent with the second-mentioned character's location than when there was no gesture (and than when the gesture went to the location previously associated with the first-mentioned character). Thus, cues to the speaker's intentions can override interpretations guided by the structural biases.

Research in children is primarily focused on understanding whether and when children's interpretation of personal pronouns is affected by the same information as adults'. The basic picture that emerges is one of knowledge developing through to the early school years. Arnold *et al.* (2007) investigated whether four- and five-year-old children prefer subjects / first-mentioned characters as antecedents, and found that children at this age showed no evidence of adult-like interpretation, measured either in off-line tasks or their eye-movements. (Children were looking at pictures showing either the first-mentioned or second-mentioned character performing a task described with a pronoun, as in example (1).) Other studies, however, have found that children are sensitive to these same cues earlier. Song and Fisher (2005), for example, found that when an antecedent was introduced first, occurred as a grammatical subject, and was the clear topic, even three-year-olds preferred it to

As this debate is not central to the work being described in this paper we will not explore these issues directly.

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

another potential antecedent. This could be because children were given more cues in Song and Fisher, or it could be because they clearly established one character as the topic, which is what young children seem to assume pronouns are used for (at least if we use their productions as a guide to their knowledge; Karmiloff-Smith, 1986). In a more recent study, Hartshorne, Nappa, and Snedeker (2015) examined the effect of order-of-mention and repeated mention (related to topic-hood) on pronoun interpretation in five-year-olds and found an effect of both, such that the children did prefer the subject / first-mentioned character even when it was not clearly mentioned more than once, although the size of the bias was not yet adult-like. The study mentioned previously by Pyykkönen *et al.* (2010) was investigating the influence of verb semantics on pronoun interpretation biases in children, not order-of-mention; however, their data are also relevant. They found that children generally preferred the subject / first-mentioned character at age three, although the effect was not particularly strong. Thus, it seems that children have at least some sensitivity to subject / first-mention as a cue by three years of age, but that it is very weak until some time after five years of age.

In the studies just described, subject-hood and first-mentioned status were, as is typical, intertwined, but one study has looked at just first-mentioned status in children. In an off-line task using stimuli like that in example (2), Goodrich Smith and Hudson Kam (2015) separated order-of-mention from subject-hood. They found that children show adult-like responding by seven – that is, showing a weak bias to interpret ambiguous pronouns as referring to the first-mentioned character – but no evidence of any sensitivity to order-of-mention by itself at five.

In terms of non-structural cues, children do not seem to interpret stress until quite late. Kertoy (1991) compared children's comprehension of contrastive stress and topic continuity, and found that even at fifth grade (around nine to eleven years of age) children were not shifting their interpretation based on stress on the pronoun. Given that stress indicates that the interpretation is not what it would normally be, late acquisition of stress cues to pronoun interpretation are not surprising; children first have to understand what the 'typical' interpretation would be to know what to shift away from. But children are capable of interpreting other cues to speaker intentions earlier. Goodrich Smith and Hudson Kam (2015), for instance, found that children were affected by co-referential gestures in ways similar to adults by seven years of age. Yow (2015) found that bilingual, but not monolingual, children will use gestures to guide pronoun interpretation by four years of age, which is earlier than Goodrich Smith and Hudson Kam (2015). She suggests that bilingual children show earlier use of the gestures than monolinguals because they are more used to monitoring a speaker's verbal and non-verbal cues to figure out meaning in context, and so are more sensitive to gesture. Note, however, that in her study the gestures were towards pictures of the characters, and so more like deictics or points than the co-referential spatial gestures used by Goodrich Smith and Hudson Kam, making direct comparisons difficult. In summary, adults use a variety of structural and non-structural cues to interpret pronouns, and the evidence shows that such sensitivity to such cues develops over time in children, which suggests that children must learn to use cues to pronoun interpretation.

The current study

We ask whether we can induce a first-mentioned bias in children's pronoun interpretations through exposure to narratives in which ambiguous pronouns refer to

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

the first-mentioned entity but the intended referents were made clear through the speaker's gestures. We conducted the study with five-year-olds, as prior work has shown they do not yet have this bias in their interpretations (Goodrich Smith & Hudson Kam, 2015). We used co-referential localizing gestures as the cue to the speaker's intended message, enabling the children to interpret the otherwise ambiguous pronoun. This is important; if the child cannot figure out that the pronoun is supposed to refer to the first-mentioned character, there is no pattern present in their experience of the input to infer.

We chose to use co-speech gesture as the cue to intended meaning for several reasons. First, unlike stress, it does not require that children already know the basic pattern; a cue that means something like 'switch-reference' is only useful after the regular pattern has been acquired, making it useless for learning from. Second, unlike eye-gaze or pointing, it can be used whether the intended referents are present or not. Thus, localizing gestures can potentially serve to help learners figure out meanings in the absence of knowledge of structural cues and they can be used for any referent, present or absent. Third, we know that children can interpret these sorts of gestures, which is a precondition for this sort of training to work (Goodrich Smith & Hudson Kam, 2015; Yow, 2015). Finally, there is evidence from a variety of studies showing that children can learn both specific things and more abstract information from gesture. For instance, iconic gestures can help children learn the meanings of novel nouns (Capone & McGregor, 2005) and verbs (Goodrich & Hudson Kam, 2009; Mumford & Kita, 2014). One study even showed that gestures led to better learning, where better learning meant more general and therefore appropriate meanings, than performing the action labelled by a novel verb (Wakefield, Hall, James, & Goldin-Meadow, 2018). This is not the only study to show generalization or abstraction of the information in gesture, however, and several studies show that gesture can lead to generalization of concepts that are more abstract than the meaning of a word. This is relevant because the generalization we are investigating is of this nature; it is not specific to a particular word, but rather, a pattern that applies over a set of words of a particular type. For instance, Cook, Duffy, and Fern (2013) found that second- to fourth-graders were better able to understand the concept of mathematical equivalence if the teacher gestured towards parts of the equation while explaining the concept, where understand means solve novel problems and even novel problem types. They found, for example, that third-graders who were trained on problems like $4 + 5 + 6 = __ + 6$ were better able to solve problems like $2 \times 4 \times 3 = __ \times 3$ when the teacher gestured while she explained the problems during training, as compared to children who saw the same teacher not gesturing.

Thus, gesture can provide the specific information we need children to be able to understand in our study, and has been shown to lead to gains in new knowledge (beyond specific exemplars), making it ideal for use in this study.

In addition, when thinking about training, we were mindful of the potential for use beyond our study. In particular, we wanted to investigate an intervention that could be adapted for use in a clinical setting, and there is reason to believe that gesture may be a particularly useful intervention for children with language delays. Research by Botting, Riches, Gaynor, and Morgan (2010), for example, suggests that children with Specific Language Impairment (SLI) might be more likely than typically developing children to rely on gesture. And recent evidence shows that, at least for word learning, gesture can improve learning in children with SLI (Vogt & Kauschke, 2017), and that this

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

benefit persists over several weeks (Lüke & Ritterfeld, 2014). Thus, studies suggest that children with language impairment not only perceive information presented in gesture, but also that gesture may serve as a compensatory device, and therefore a potentially useful clinical intervention. Moreover, studies show no advantage of naturalistically collected gestures over systematically produced ones when gesture's impact on learning is measured (Hostetter, 2011), meaning that training people to produce certain kinds of gestures is a reasonable and potentially powerful clinical intervention.

Related, one might ask whether children actually see localizing gestures in their input, and so question the relevance of the cue outside of the laboratory. On the one hand, there is evidence that speakers, even children themselves (So & Wong, 2018), do produce such gestures, although not very frequently (So, Coppola, Licciardello, & Goldin-Meadow, 2005; So *et al.*, 2009). Thus, while infrequent, such gestures likely do occur in children's naturalistic input, and so could potentially be involved in learning processing preferences in the wild. On the other hand, our study was intended more as a demonstration of possibility. We were not trying to show that children actually learn a first-mentioned bias from gestures naturalistically (that would require a different sort of investigation). Rather, we were interested in whether we could induce a structural bias in children that they didn't exhibit prior to training. If so, it would provide stronger evidence that such biases might indeed be learned via patterns in the input.

Note that we are not interested specifically in the first-mentioned bias per se, but it has several properties that make it useful in a study such as ours. One, it is real. Adults show this bias (Goodrich Smith & Hudson Kam, 2012), as do children by age seven (Goodrich Smith & Hudson Kam, 2015), so inducing this bias in children would not harm their use of language in the real world, it would simply lead to children having the processing bias earlier than they would normally. Two, it is fairly clear that this cue has no inherent grammatical, communicative, or functional source or reason; it happens to be the case that pronouns tend to refer to the first-mentioned entity, but not always, and it need not be so. It is a reflex of other aspects of language. For example, topics tend to be old information, old information tends to be in subject position, and subjects precede objects in canonical sentences in most languages; thus, if a pronoun refers to a topic it will tend to refer to the first rather than the second entity. Despite this epiphenomenal origin, there is evidence that listeners eventually pick up on this tendency to the point where it (weakly) affects pronoun interpretation (Goodrich Smith & Hudson Kam, 2012, 2015). Gernsbacher and Hargreaves (1988, 1992), propose that this tendency arises from general cognitive principles – the well-known primacy effect; however, they did not have direct evidence for this in pronoun interpretation. If the first-mention bias really was an effect of primacy, one might expect it to be evident even in the youngest children, which is not the case. Thus, it is hard to make the argument that such effects are due to anything other than having implicitly learned the tendencies in how pronouns are linked with antecedents. And finally, there is the fact that order-of-mention cannot be removed from any stimulus in which multiple characters are mentioned. Due to the sequential nature of speech, one character will necessarily be mentioned before any other(s). Thus, it is a cue that cannot be removed, that we know people learn eventually, and that has little to no communicative, grammatical, or functional reason. Together these things make order-of-mention a useful test case for a study examining the learning of processing biases.

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

Comparison conditions

If we find that children learn this processing bias from gesture, there is a variety of possible explanations as to why. On the one hand, it could be that gestures are particularly effective at producing generalizations. Alternatively, it could be that the ambiguity present in the speech was important, leading the children to simultaneously look for ways to interpret the ambiguous pronoun (to be found in the gestures) and pay attention to the structures themselves. Valian and Casey (2003) suggest that successful language interventions (that is, ones that lead to generalizations) focus the child's attention on syntax rather than meaning. Although the stimuli in our gesture condition do not focus the children's attention on syntactic structure specifically, it does require that they encode (and therefore presumably attend to) something beyond the semantics of the words in the sentences, and thus might be particularly effective in helping children induce the first-mentioned tendency. If so, that is, if the ambiguity leads children to focus on structure and it is the focus on structure that produces a generalization, then we shouldn't see learning of the bias when children do not have to deal with ambiguity. We contrast these two possibilities in a second condition: the Gender Training condition. Children in this condition heard narratives with two differently gendered characters, so it was always clear which of two characters were being referred to by the pronoun. As in the Gesture Training condition, the character introduced first was the one later referred to with the pronoun. Thus, these narrations also exhibited the first-mentioned tendency; however, the semantics of the words themselves clearly disambiguated the antecedent, unlike the narrations in the experimental condition. In a third condition, the Name Training condition, we further sought to understand how specific any learned pattern might be. Specifically, does the first-mentioned pattern need to be instantiated by pronouns for it to be learned, or is it something about reference more broadly? The training narratives these children heard did not contain pronouns. Instead, character names were repeated. Importantly, these latter two conditions allow us to better understand any effect we might see in the primary condition of interest (the gesture condition).

Training and language development

A necessary precondition for the current study, that children's performance will improve with training, is supported by numerous studies demonstrating that children can learn various aspects of language earlier than they might otherwise through repeated exposure in a lab setting. For instance, Nelson (1977) gave two groups of 2½-year-old children either extra exposure to *yes/no* question constructions or complex verb phrase constructions. Each group of children learned the construction they were exposed to earlier than the other group, which was also earlier than they normally would. That is, each group's development was accelerated only for the construction they were exposed to. More recently, Valian and Casey (2003) were targeting inversion in *yes/no* questions and found that repeated concentrated exposure to *yes/no* questions with inverted auxiliaries improved performance on a repetition task in children aged 2;6–3;2. And several studies have shown that with as few as two 30-minute training sessions, children as young as three years old can produce a full passive utterance with a novel verb (e.g., *The elephant was frimmed by the dog*), as much as one year earlier than this construction is typically produced by

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

children (Brooks & Tomasello, 1999; Pinker, Lebeaux, & Frost, 1987; Tomasello, Brooks, & Stern, 1998).

Even younger children have been found to benefit from somewhat longer training studies. In a 9-week training study, Smith, Jones, Landau, Gershkoff-Stowe, and Samuelson (2002) exposed 17-month-olds to names for unfamiliar objects repeatedly. The labels referred to categories organized by shape. At the end of the training the infants not only extended the trained names to new exemplars with the same shape as the training exemplars, they also showed a shape bias for novel-at-test names and objects earlier than is typical. On the basis of the exposure in the study the infants had formed the generalization that objects with similar shapes have the same name. Moreover, and of particular relevance to the current work, the trained infants showed an increase in their acquisition for new object names outside of the study, as compared to the control infants. Thus, even very young children can 'learn to learn' through a training study, suggesting that the children taking part in the current work could also move on from relying on gesture as a cue to pronoun reference, and make a generalization about anaphoric pronouns more generally.

Studies have shown that processing biases can be manipulated via mere exposure as well, in adults at least. Wells, Christiansen, Race, Acheson, and MacDonald (2009) exposed adult native-English speakers to sentences of a type known to be quite difficult to process (object relative clauses) and found that participants became much faster at reading them, suggesting they became less difficult to process. Thus, there is evidence that training is not just applicable to particular structures, but can potentially also affect aspects of language processing.

Method

Participants

A total of 50 five- to six-year-old children from local preschools, kindergarten classrooms, and daycare centres took part (range: 5;0–6;4, mean: 5;6). (The target age was five years. Some of our participants were recruited and run in kindergarten classrooms, so that sometimes they were young six-year-olds. Age information was not provided for two of the children who did not complete the study.) Among them were 21 Caucasian children, 9 African American children, 3 East Asian children, 1 Hispanic child, and two who were mixed race according to parental report. The parents of the remaining six children did not provide us with ethnicity information. The consent form specified that participating children should be native English speakers and explained that that meant that "English is the primary language spoken in your home, although it is fine if your child is also exposed to other languages". In addition, we obtained information about the home language from 37 of the families. Of these, five were also exposed to a language other than English in the home. For three children, this non-English exposure was minimal. For two of the children, the other language was used quite regularly, although not as much as English. Eight children did not complete the entire study, leaving 21 female and 21 male children in the final sample.

The children were randomly assigned to one of three conditions (Gesture Training, Gender Training, and Name Training), such that fourteen children were in each condition. Demographic information for the children who completed the study is shown by condition in Table 1.

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

Table 1. Participant age and gender by condition

	Condition		
	Gesture	Gender	Name
Age (mos): Mean (SD) (Min, Max)	68.07 (1.17) (62, 76)	64.5 (0.948) (60, 73)	64.93 (1.211) (60, 75)
Gender (N female/total N)	9/14	5/14	7/14
Average number of days to complete the study (Min, Max)	13.21 (4, 22)	7.71 (5, 14)	12.92 (5, 22)
Average training sessions adjacent to testing	2	1.5	2.17

Pre- and Post-training test

The test was the 'No Gesture' video used in Goodrich Smith and Hudson Kam (2015). It contained eight videotaped narrations. Five of the narrations contained ambiguous pronouns referring to the characters; these five narratives were the test items. Children answered questions about the events in the narrative after each one, including a question regarding the referent of the pronoun. These questions established each child's individual propensity to use order-of-mention as a cue when interpreting pronouns. Three distracter narrations that did not probe ambiguous pronouns were also included. These took place at the beginning, middle, and end of the test session to allow the child to feel confident in their performance. Crucially, the narrator did not produce localizing gestures while speaking in the test video. (The same video was used for pre- and post-testing.) Example testing and filler narratives, including the questions asked after each, are presented in Table 2, and the complete set is presented in the 'Appendix'.

Participants' responses to the items asking about the referents of the pronouns were coded according to whether the child responded with the first-mentioned character or not, and the results from the pre-test were compared to those from the final session to assess the impact of training. Sometimes a child did not answer the question, or provided an uninformative response (e.g., "I don't know"). In these instances, the child was prompted by the experimenter to provide a name. Specifically, if the child said they didn't know, the experimenter said, "It's okay, you can just guess", and if they still needed help she said "Do you remember if it was X or Y?" In these instances the choices were always presented (in the experimenter's question) in the same order as the narrative. When this happened the coding included the fact that it was a prompted response.

Training and training conditions

The study included three days of training. In each training session the children watched a video containing 10 new narrations (narrations that were distinct from the pre-test narrations and from each of the other training days' narrations). After each narration, the experimenter asked the child which character performed a given action in the story they had just heard.

The training narrations differed slightly between the three training conditions. In the Gesture Training condition, children watched videos similar to the order-of-mention condition in Goodrich Smith and Hudson Kam (2015); the narrations contained two names of the same gender and an ambiguous pronoun that followed later in the

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

Table 2. Sample test items

Narrative type	Narrative and questions
Ambiguous Pronoun Narrative	Donald and Mickey are good friends. Donald is walking up a hill, and Mickey is at the top. He has an umbrella, which is good because it looks like it's going to rain.
	Who has an umbrella?
	Are Donald and Mickey friends?
	What is the weather like?
Filler Narrative	It's a warm sunny day. Billy and Sam are playing catch with a baseball. Billy isn't very good at catch because he drops the ball a lot.
	What is the weather like?
	Who is playing catch?
	Is Billy good at playing catch?

narration. The narrator produced localizing gestures with each name, and while saying the pronoun gestured in the location previously matched with the first-mentioned name. Example gestures are shown in Figure 1. (These are still captures from the video; the gestures in the videos were dynamic.) The subsequent questions, asked live by the experimenter, inquired about the referent of the ambiguous pronoun.

The Gender Training condition used a similar script, the only difference being that the gender of the two characters differed. In this condition the gender of the pronoun always matched the gender of the first-mentioned character. The narrator did not produce gestures while speaking in this condition. In this condition the question after the narrative asked about the referent of the unambiguous pronoun. The Name Training condition had the same script as the Gesture Training condition, except that, rather than saying a pronoun, the narrator said the first character's name again. As in the Gender Training condition, the narrator did not gesture while speaking. Thus, the children in the Name Training condition were also exposed to a first-mentioned pattern of reference (the first-mentioned character is referred to again), but not with respect to pronouns. In this condition the question after the narration asked about the repeated name. Note that the actual question was the same across the three conditions, what differed was how the critical referent had been referred to in the description of that activity in the narrative. Table 3 shows an example narration as it occurred in each of the three conditions.

Note that the training narratives in all three conditions contained a first-mentioned pattern of reference (the first-mentioned character is referred back to in each narrative). Thus, it was possible that children in all three conditions would pick up on the pattern and so be more likely to select the first-mentioned name as the referent for the ambiguous pronoun after training than before. However, we anticipated that children in the Gesture Training condition would show this effect most strongly because they received input that was most relevant to interpreting ambiguous pronouns, which is what we were testing.

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press



Figure 1. Example gestures from the Gesture condition narratives: (A) Localizing Gesture with first-mentioned name; (B) Localizing Gesture with second-mentioned name; and (C) (first-mentioned) Co-referential. (Copied from Goodrich Smith & Hudson Kam, 2015).

Table 3. Example narrations from the three training conditions

Condition	Example narration
Gesture Training	My mom and my sister are working in the garden. My mom is planting flowers, and my sister is planting vegetables. She really hates pulling weeds.
Gender Training	My mom and my brother are working in the garden. My mom is planting flowers, and my brother is planting vegetables. She really hates pulling weeds.
Name Training	My mom and my sister are working in the garden. My mom is planting flowers, and my sister is planting vegetables. My mom really hates pulling weeds

Procedure

On Day 1 of the study all children took part in the pretest. On Days 2, 3, and 4 of the study children watched the training video associated with their condition. On the final (test) day, all children, regardless of condition, watched the video they had seen at pre-test and answered questions regarding the referent of the pronoun. Recall that the only potential cue to pronoun resolution in these stories was order-of-mention. This post-test served to measure whether the training had any impact by assessing whether their interpretation of ambiguous pronouns had changed over the course of the study.

Children were tested at their kindergarten (N = 24) or preschool or daycare (N = 18) in the Berkeley/Oakland area, either in a hallway outside of the classroom, or in a separate room near the classroom. The study took five sessions, which typically occurred over the course of two school weeks although it sometimes took place over three. The three-week spans generally occurred because the first session, in which the children's baseline interpretation tendencies were assessed, took place in a different week than the training and post-training testing (which sometimes took one week and other times was spread over two). The specific arrangement of days varied by scheduling logistics at each school and the child's individual attendance. The rationale for multiple sessions spanning several days rather than a single training session was as follows: we assumed, based on previous studies discussed earlier, that multiple training sessions would be necessary for children to show an improved performance, especially given the role of sleep in memory enhancement and consolidation (Maquet, Peigneux, Laureys, & Smith 2002; Stickgold & Walker, 2005).

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

For instance, Gómez, Bootzin, and Nadel (2006) found that 15-month-old babies who napped after learning an artificial language performed better at abstract rule generalization than non-napping infants. These findings suggest that children will show more learning, and possibly more generalization – an important component of this study – if given time to sleep in between sessions.

The average number of days taken to complete the study was 11.2 (including weekend days in the count; range 4–22). Note that for all of the children who took 22 days to complete the study, the pre-test took place more than 2 weeks before the rest of the study, which was completed within a single school week. The average number of school weeks that it took the children to complete the training and post-test (so, ignoring the pre-test) was 1.625, with a range of 1–2 weeks. For most children, at least one training session occurred immediately prior to the post-test, and the average training days immediately prior to testing was 1.875. Fourteen children had all three training sessions in the days immediately prior to the post-test, 10 had 2, 13 had 1, and 3 had zero, that is, testing occurred in a different school week than training. (The information on training dates was unfortunately not recorded for two of the children.) Average training information by condition is presented in Table 1. This kind of spread is more or less unavoidable for a multi-day training study done outside the lab, where the researcher does not have a high degree of control over schedules. One side effect of this, however, is that it means that any findings are likely to be more generalizable; they are unlikely to depend on a highly specific set of circumstances as the children got varied training and test schedules.

Results

Pre-test

Figure 2 shows the pre-test and post-test scores for the three conditions. Looking first at the pre-test (prior to training), the average percent of first-mentioned responses for the Gesture Training condition was 51.43%, for the Gender Training, 52.86%, and for the Name Training, 47.14%. Thus, they all hovered around chance, suggesting that the children were not showing a response bias of any kind. The overall rates of first-mentioned responding pre-training also fit well with the performance of the five-year-old children in the No Gesture condition of Goodrich Smith and Hudson Kam (2015), who selected the first-mentioned character 48% of the time.

The effect of training

The mean rates of first-mentioned responses went up in all three conditions, but they went up most in the Gesture Training condition (to: Gesture Training – 69.05%; Gender Training – 57.14%; Name Training – 55.71%; increases of: Gesture Training – 17.62%; Gender Training – 4.29%; Name Training – 8.57%).

We examined the impact of training on children's choice of first- or second-mentioned character by the different training conditions using mixed effects logistic regression (R Version 3.3.3; R Core Team, 2017; lme4 package: Bates, Maechler, Bolker, & Walker, 2015). To account for idiosyncratic within-subject tendencies, we included random intercepts and slopes for subjects pre- and post-training. As the model involves a comparison between the three training conditions, we ran three

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

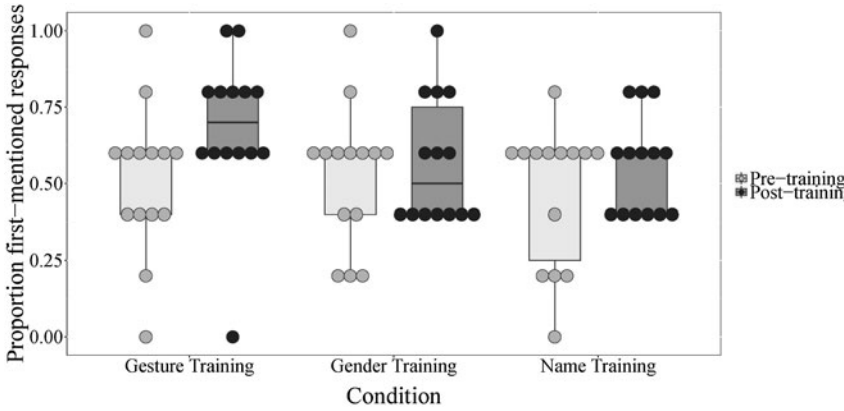


Figure 2. Proportion first-mentioned responses pre- and post-training by condition. (In this and in Figure 3, the lower and upper hinges on the boxes correspond to the first and third quartiles (the 25th and 75th percentiles). Whiskers extend to the highest (or lowest) actual data point no further than 1.5 * the inter-quartile range away from the hinge. When they occur, dark solid lines inside the boxes represent the median.)

iterations of the model with each condition as reference. These results are presented in Table 4. Though it is clear from Figure 2 that children in the Gesture Training condition showed larger gains from training than those in the other two conditions, this difference did not significantly differ between conditions (as indicated by the interaction terms). The increase in choice of first-mention post-training was not significant in either the Gender or Name training conditions; however, the increase in first-mention choice in the Gesture Training condition was significant (OR = 2.10, 95% CI = [1.04–4.24], $p = .039$).

A potential reason for the lack of a significant interaction (which would demonstrate a difference in the effectiveness of training between the three conditions) is greater heterogeneity of response to the training in some conditions as compared to others. Numerically, children in the Gesture Training condition were more likely to increase first-mentioned responses than they were to decrease or stay the same, something not true of children in the other two conditions, as shown in Table 5. That is, Gesture Training had a more consistent effect. A Pearson chi-square analysis on these data was not significant ($\chi^2(4, N = 42) = 3.90$, $p = .420$), however. In a second analysis we removed the children who showed no change and compared the number of children who increased versus decreased in their first-mentioned responses with training. We did this because, if responses are truly random, the likelihood of no change is actually much smaller than the likelihood of change. This analysis was not significant either ($\chi^2(2, N = 30) = 3.90$, $p = .142$).

Next we looked at change (direction and occurrence) dependent on starting point. We looked at whether the child started above or below 50% first-mentioned responses, how that affected whether they increased, decreased, or stayed the same (in terms of first-mentioned responses), and whether condition interacted with this. The rationale behind this analysis is as follows: if a child's starting point is due to chance, and responding post-training is also due to chance, then children who were showing a bias pre-training should be just as likely to not show a bias post-training as to show one, or perhaps, given that they have more room to move down than up, even less likely to show the bias. But if there is an effect of training, we would expect

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

Table 4. Model results for choice by training session and condition

Model 1. Response ~ Training * Condition + (Training | Subject)									
Fixed effects									
	Ref = Gesture			*Ref = Gender*			*Ref = Name*		
	Odds Ratio	*95% CI*	*P*	*Odds Ratio*	*95% CI*	*p*	*Odds Ratio*	*95% CI*	*p*
(Intercept)	1.06	0.63–1.79	.824	1.13	0.67–1.90	.655	0.89	0.53–1.50	.655
Gesture				0.94	0.49–2.25	.903	1.19	0.57–2.50	.636
Gender	1.06	0.51–2.22	.874				1.27	0.61–2.65	.528
Name	0.84	0.40–1.75	.636	0.79	0.38–1.65	.528			
Training	2.10	1.04–4.24	**.039**	1.19	0.60–2.35	.613	1.42	0.72–2.80	.306
Training : Gesture				1.76	0.66–4.67	.257	1.47	0.55–3.91	.487
Training : Gender	0.57	0.21–1.51	.257				0.84	0.32–2.18	.716
Training : Name	0.68	0.26–1.80	.437	1.20	0.46–3.12	.716			
Random effects									
$\tau_{00,\ Subject}$					0.16				
ρ_{01}					−1.000				
$N_{Subject}$					42				
$ICC_{Subject}$					0.046				
Observations					420				
Deviance					546.94				

Table 5. Number of children in each training condition who increased, decreased, or showed no change in their first-mentioned responses with training

	Response to training		
Condition	Decrease	No change	Increase
Gesture Training	1	4	9
Gender Training	5	4	5
Name Training	4	4	6

children to increase their first-mentioned responses regardless of where they started. Table 6 shows the data broken up this way. The chi-square analysis (starting over vs. under by change with condition as a layer) shows that whether a child starts over or under has a significant impact on whether and how they change ($\chi^2(2,\ N = 42) = 11.823$, $p = .003$). This effect is not the same for all three conditions, however, it is only significant for children in the Gender condition (Gesture: $\chi^2(2,\ N = 14) = 4.926$, $p = .085$; Gender: $\chi^2(2, N = 14) = 7.249$, $p = .027$; Name: $\chi^2(2, N = 14) = 4.926$, $p = .085$).

The previous analysis assumes that all children are equally able to increase or decrease their first-mentioned responses, which is not necessarily the case, so in one final contingency analysis we removed children who were at the floor or ceiling pre-training, resulting in four children being excluded from the analysis: two who were at ceiling and two who were at the floor. The chi-square results were: overall: $\chi 2(2,\ N = 38) = 11.588$, $p = .003$; Gesture: $\chi 2(1,\ N = 12) = 3.592$, $p = .083$; Gender: $\chi 2(2,\ N = 13) = 6.451$, $p = .04$; Name: $\chi 2(2,\ N = 13) = 3.846$, $p = .146$). It remains the case that, while the overall starting point affected how children changed, this was only significant in the Gender Training condition.

These last two chi-square analyses suggest that changes in the Gender Training condition show a larger effect of chance than changes in the other two conditions, while changes in the Gesture and Name Training conditions reflect learning more than those in the Gender Training condition. Note, however, these contingency analyses were post-hoc and done with small samples; the results should therefore be viewed with caution.

Recall that we also recorded whether a response was prompted, and it is possible that the effects of training and condition might be revealed through the relative confidence of a child's response in addition to their actual answer; therefore, we ran a logistic mixed effects model predicting children's need for prompting by condition and training. Model results are reported in Table 7. As can be seen there and in Figure 3, children were less likely to require a prompt after training in the Gesture training (OR = 0.26, 95% CI = [0.08–0.84], $p = .025$) and Name training (OR = 0.17, 95% CI = [0.04–0.61], $p = .007$) conditions. While prompting did not significantly differ by training session in the Gender training condition (OR = 0.79, 95% CI = [0.24–2.62], $p = .705$), there was no significant interaction between condition and training in any version of the model.

Discussion

This work explores the learning of a processing bias in pronoun interpretation by children. Specifically, we examined the first-mentioned bias, which is a bias to

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

Table 6. Number of children in each training condition who increased, decreased, or showed no change in their first-mentioned responses with training, by initial response pattern

	Response to training		
Condition	Decrease	No change	Increase
Gesture Training			
Over 50%	0	4	4
Under 50%	1	0	5
Gender Training			
Over 50%	5	3	1
Under 50%	0	1	4
Name Training			
Over 50%	4	3	2
Under 50%	0	1	4

interpret pronouns as the first-mentioned character when no other cues to the likely referent are present. We investigated whether we could induce this bias in children with training comprising input that evinced the pattern. We chose this particular processing bias because it is real, as in, it exists as a bias in adults and older children (Goodrich Smith & Hudson Kam, 2012, 2015), and so inducing this bias in children would not harm their use of language in the real world. And, as it is not derived from other potentially innate biases, studying the emergence of this bias with training has the potential to inform our understanding of how other biases might be learned.

Training in the primary condition consisted of repeated exposure to stories that contained two characters of the same gender introduced in syntactically parallel ways, so that neither character was syntactically or pragmatically more salient than the other. A later sentence described a single character as engaging in some other activity using a third-person singular pronoun. The referent of the pronoun was ambiguous based on the speech alone if the listener had not yet acquired the first-mentioned bias. In the videos in this condition the speaker gestured as she spoke in ways that indicated her intended referent was the first-mentioned character. If the child could interpret the gestures they could interpret the pronouns and therefore potentially induce a first-mentioned bias for ambiguous pronouns. The data show some evidence that children did learn the processing bias in this condition, supporting the idea that such processing biases are learned, and that gesture can serve as a useful cue to interpretation in training.

Note our definition of the first-mentioned bias: a bias to interpret pronouns as the first-mentioned character when no other cues to the likely referent are present. Given this, we started with the assumption that ambiguity was important for the learning of this processing bias. Gesture provided the resolution of the ambiguity. But note that, unlike gestures used in many other studies showing beneficial effects of gesture on learning (Hostetter, 2011), the gestures we used were not directly iconic nor were they actually spatial. They required some interpretation: the child had to track the initial locations (or sides) associated with each referent and remember the

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

Table 7. Model results for prompted/not prompted by training session and condition

Model: prompted/not prompted ~ condition * training + (training | subject)									
	Ref = Gestures			***Ref = Gender***			***Ref = Name***		
Fixed effects									
	OR	*CI*	*p*	*OR*	*CI*	*p*	*OR*	*CI*	*p*
(Intercept)	0.61	0.28–1.32	.208	0.29	0.12–0.68	**.004**	0.49	0.22–1.07	.074
Gesture				2.09	0.67–6.53	.207	1.24	0.41–3.73	.703
Gender	0.48	0.15–1.50	.207						
Name	0.81	0.27–2.43	.703	1.68	0.54–5.27	.371	0.59	0.19–1.86	.371
Training	0.26	0.08–0.84	**.025**	0.79	0.24–2.62	.705	0.17	0.04–0.61	**.007**
Training : Gesture				0.32	0.06–1.62	.170	1.56	0.29–8.33	.606
Training : Gender	3.09	0.62–15.48	.170				4.81	0.89–25.86	.067
Training : Name	0.64	0.12–3.44	.606	0.21	0.04–1.12	.067			
Random Effects									
$\tau_{00,\ Subject}$					1.136				
ρ_{01}					−0.725				
$N_{Subject}$					42				
$ICC_{Subject}$					0.257				
Observations					420				
Deviance					354.626				

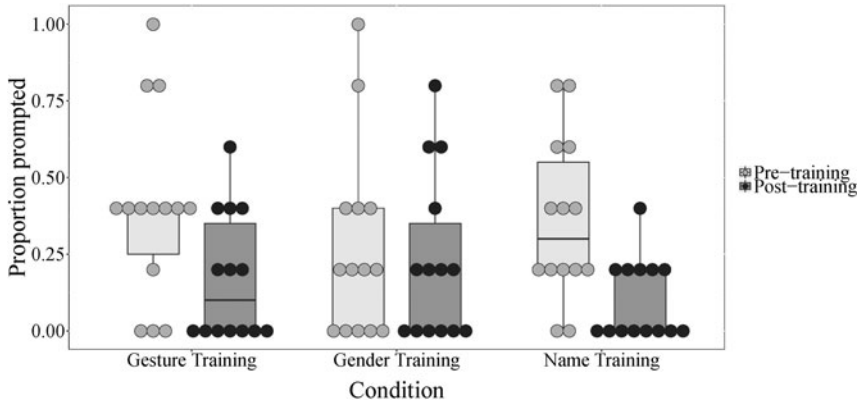


Figure 3. Proportion prompted responses pre- and post-training by condition.

associations in order to use the gestures. We assumed, based on previous work (Valian & Casey, 2003), that this difficulty may have been crucial to the generalization.

We included another condition to assess whether ambiguity was actually necessary for the first-mentioned bias to be extracted. In this condition the narrations were exactly the same except that the two characters were of different genders, and so the referent of the pronoun was clear in context. We reasoned that the ease of interpretation based on purely semantic features may have allowed the children to avoid attending to the structure of the sentences and thus result in the children in the Gender Training condition failing to pick up on the first-mentioned pattern. The third condition asked whether the first-mentioned bias was just about pronouns. In this condition all references to the characters involved names. If the first-mentioned bias is about pronouns specifically, we should not see children in the Name Training condition develop a first-mentioned bias with training.

The overall data show some evidence for children in these two conditions having learned a first-mentioned bias as well, although unlike the Gesture Training condition, the likelihood of first-mentioned responses did not increase significantly in the Gender and Name Training conditions. However, the magnitude of change between the three conditions did not significantly differ either, making it difficult to make any strong claims about the two additional conditions.

Nevertheless, the way they pattern with respect to each other is interesting. Numerically, while the Gesture Training condition showed the largest increase in first-mentioned responses with training, the Name Training condition was the next most effective, not the Gender Training condition, this despite the fact that the former did not involve pronouns while the latter did. Why this is so is not clear. It may be the case that, because the Gender Training condition narrations contained a cue to the referent of the pronoun that is well controlled by children at this age (Arnold *et al.*, 2007), they had little reason to look for or notice another cue. This line of reasoning is consistent with Arnold *et al.*, who point out that, due to their more limited language processing recourses, children might be more likely to focus their attention on highly reliable cues (like gender), while paying less attention to less reliable cues such as order-of-mention, particularly when they are trying to bring multiple sources of information together in real-time processing. Or it could be a

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

blocking effect. Blocking is a classic and widespread phenomenon in learning that is basically the following: when a new pattern or information source is not informative above and beyond what is already known, it is not learned. (See Ramscar, 2013, for a discussion of blocking and its relevance to language learning.) In the case of the stimuli in our study, because gender is known, order-of-mention is not learned because it is not more informative.

But it is not the relatively small or even lack of an effect of training in the Gender Training condition we found the most surprising, it was the performance in the Name Training condition: numerous children in the condition with no pronouns in training also apparently learned the first-mentioned pattern, and moreover, transferred it to pronouns. This numerical pattern was paralleled in the prompting data, which reflects children's relative level of certainty on the task: children in the Gesture and Name Training conditions were less likely to require prompting after training, while children in the Gender condition did not significantly change. Perhaps this apparent learning was due to the oddity of repeated names. Generally, we do not repeat names, especially three times in a short narrative, unless we are highlighting the name in some way (Arnold, 2010). And the repeated names penalty in processing which is exhibited by adults (Gordon, Grosz, & Gilliom, 1993) and children alike (Megherbi & Ehrlich, 2009) suggests that speakers are sensitive to its low frequency. The relative strangeness (statistically speaking) of the repeated names in the narrative may have drawn children's attention to the structure, and thereby allowed them to learn the order-of-mention pattern.

If this is indeed what occurred, it suggests that (at least) the children in the Name Training condition learned a pattern not with respect to pronouns, but with respect to reference more generally. This point is quite interesting. Arnold (2013) reminds us that pronouns are but one kind of reference, and that patterns may be shared across different kinds of reference. If the Name Training children did indeed learn a pattern from names and apply it to pronouns, then it provides evidence of the level of generalization being reference more broadly, not just pronouns. We cannot know if all children, regardless of condition, who seem to have acquired the first-mentioned bias learned something as general as those in the Name Training condition, but it is possible. To test that would require a follow-up study examining how they process other kinds of reference. For instance, one could look at how quickly they fixate on a named referent when it was the first- vs. second-mentioned previously. If children in our study had acquired a broad pattern applicable to all cases of reference, we would expect them to be faster to look towards a first-mentioned character when it is later mentioned as compared to a second-mentioned character when it is later mentioned.

The data from the exploratory contingency analyses, however, suggest caution in interpreting the Name Training data. Although, overall, there were no differences in the proportion of children who went up, down, or stayed the same in the rate of first-mentioned responses, when starting rate is considered, numerically, only the children in the Gesture Training condition were as likely to go up when they started above 50% as when they started below. Children in the Name Training condition, in contrast, were more likely to go down if they started over 50% and up if they started under, suggesting that there was a large element of chance involved in their responses before and after training. The same is true for children in the Gender Training condition. Thus, the average changes in the Name Training condition may just be more of a regression to the mean phenomenon. (Here we are discussing the numerical trends. The statistical tests were only significant for the Gender Training condition.)

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

One thing that did emerge from training across the board was increased general ability and confidence in interpreting pronouns, as measured by the rate of prompting. This was true even when training did not include pronouns. Thus, children were learning to pay attention to the narratives so that they could answer questions about them. This confidence did not interact with their interpretations, however.

Caveats

Although we think the results show learning, especially in the Gesture condition, there are some points about the study that should be acknowledged. First, the sample is small. This is almost unavoidable for a multi-day training study like this, but it has important implications. Of greatest concern is the fact that it means that the patterns of significance / lack thereof are unlikely to replicate. However, we see more importance in the direction of the results than the specifics of the statistical significance – which will always vary between samples – and our data show a larger effect in exactly the condition where it is predicted. Moreover, the overall pattern of results – that is, suggestions that a first-mentioned bias was learned – exist whether one is looking at the individual responses or the patterns (i.e., the contingency analyses), giving us some confidence that the effect is likely to be real, if small. Of relevance, the actual pure first-mentioned bias in adults and older children is itself weak (Goodrich Smith & Hudson Kam, 2012, 2015), and a weak bias would only ever be expected to be a small effect.

Second, there is the question of what exactly was learned. One reviewer raised the possibility that it was just a specific pattern associated with a specific speaker. Although we cannot rule this out, as the narratives were all produced by the same speaker, we think this is unlikely. First, there is the fact that prior training studies have found generalization, not only to untrained items, but also to language learning outside the lab (Smith *et al.*, 2002). Second, there is the fact that the testing videos were not the same as the training videos in the information they contained. If the children had truly been learning something that specific during exposure we would have expected it to be tied to the particular training stimulus. Thus, they would have shown the first-mentioned pattern in response to training videos but not to the post-test videos. In fact, we might then expect continued chance performance post-training, as they would have developed a reliance on gesture to interpret pronouns, something that they couldn't use in the tests. It is true that we used the same video in pre- and post- training testing. On the one hand, this means that participants could have developed a familiarity with that video that affected their performance. However, there is no reason that familiarity with the specific videos would lead to increased first-mentioned responses, so this is unlikely to explain the results. On the other hand, it means that any differences in pre- and post-training responses cannot be attributed to differences in the narratives tested, which is a strength of the design.

It is possible that the children were learning a speaker-specific bias, rather than a bias associated with the specifics of the experiment, but extant data suggest that that is unlikely, as speaker-specific syntactic patterns seem not to be easily learned (Hudson Kam, unpublished observations; Ryskin, Fine, & Brown-Schmidt, 2017; Ryskin, Qi, Duff, & Brown-Schmidt, 2017). Second, why would this be the case only (or more so) in the Gesture Training condition? If the children are learning speaker-specific structural patterns, shouldn't they be equally likely to learn them in all three conditions? Thus, it is likely that the children learned an actual generalization.

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

Finally, there is the point that the training and testing experiences the individual children had were varied. Due to the realities of running this kind of study, children were not all available on the exact same schedule. In some respects this is undesirable. But in others, it gives the results (the potential for) higher external validity; whatever our results, they do not appear to be confined to a highly specific set of training and testing circumstances. This gives us greater confidence that this kind of training could work more broadly.

Above, we stated that the children in the Gesture Training condition learned the order-of-mention pattern as they learned to use the localizing gestures, that is, that they were learning to use the gestures during this study. However, we have no independent evidence that they did actually learn to interpret the gestures. It may, in fact, be the case that only some children learned to read the gestures, and that that variation explains the variation in the acquisition of the first-mention bias. While we cannot tell whether the children really learned to read the gestures, we can look at performance in the training and its relationship to performance on the pre- and post-training tests. Looking first at the six children who either stayed the same or decreased their first-mention responses from pre- to post-test, only one had more first-mentioned responses during training than during testing; the other five all decreased, three a great deal. That is, they were 'worse' during training than during testing. For the eight children whose rate of first-mentioned responses was higher after than before training, six produced more first-mentioned responses during training than before it, and the two who decreased during training (as compared to before training) did so only slightly. This is purely descriptive, but it does again indicate that the gestures were instrumental to learning.

Broader implications

What have we learned about pronoun interpretation/learning more broadly from this study, especially, or additionally, with respect to training? If we accept that at least some children in this study acquired a first-mentioned interpretation bias, and that this was due to the training they received, it supports the idea that processing biases more generally might be learned. Of course, we investigated a bias that seemed a more clear-cut example of a learned bias than some others might be (e.g., a subject bias), and so it still might be possible that other biases are not learned. We would point out, however, that the subject bias, if it exists in young children, is weak, and that it takes time to look adult-like. Moreover, even the first-mentioned bias has been proposed to be innate rather than learned (innate in the sense of a product of general human information processing, not as in part of UG; Gernsbacher & Hargreaves, 1992).

There is the question of what, exactly, the children in our study learned; specifically, whether they learned a processing bias that was relevant only to pronouns or was broader. The data from the Name Training condition suggest that at least some of the children may have learned a more general processing bias that applies to reference more broadly, not just to pronouns. Unfortunately, we did not test generalization in any of the children and so cannot comment on whether this is true for children in the other conditions. Moreover, the data are somewhat equivocal as to whether the children in the Name Training condition really did acquire a bias. Some of the statistical tests suggest they did, whereas others suggest they did not.

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

Then there is the question of how this kind of training might be applied more broadly. On the one hand, the data show that training is effective for the learning of processing biases, and so typical pronoun interpretation could likely be trained. There is the question of what training should look like to be most effective. The data suggest that children in the Gesture Training condition were more affected by their training than those in the other two conditions, with some statistical tests showing this, and others showing no difference between the conditions. These results might seem to some to show that gesture itself is highly potent and so interventions should include gesture. However, there are reasons to be more cautious in extending our results in this way. First, the presence of gesture was not the only difference between the Gesture Training condition and the other two conditions; the narratives themselves were ambiguous in the Gesture condition. Indeed, we assumed that ambiguity itself would drive children to learn the structural pattern present in the pronominal reference. The gestures were included so that the participants could interpret the pronouns in ways consistent with the first-mentioned pattern. Thus, it is not clear from our data that gesture is more effective than other cues that would likewise disambiguate otherwise ambiguous pronouns. It could be the case that gestures drive learning, and would do so when provided alongside unambiguous pronouns, e.g., gendered pronouns, or it could be the case that ambiguity is necessary for better learning. There are other studies showing that gesture is not always effective at helping children learn (Yeo, Ledesma, Nathan, Alibali, & Church, 2017), so it is not a given that it was the gestures per se that contributed to the learning we saw. More research is required to distinguish these two possibilities, and to guide best practices for interventions. What we can conclude based on our results is that gesture can be used to resolve the ambiguity sometimes inherent in pronouns, and that children can learn processing biases from input in which gesture serves this function.

Acknowledgements. This research was supported by National Institutes of Health Grant HD 048572 and by a Standard Research Grant from the Social Sciences and Humanities Research Council of Canada, both to CHK. The authors wish to thank the members of the Language and Learning Lab at the University of California, Berkeley (where this research was conducted) for their advice and comments on this work, and to members of the Language and Learning Lab at the University of British Columbia for comments on an earlier draft.

References

Akmajian, A., & Jackendoff, R. (1970). Coreferentiality and stress. *Linguistic Inquiry*, *1*, 124–6.

Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes*, *31*, 137–62.

Arnold, J. E. (2010). How speakers refer: the role of accessibility. *Language and Linguistics Compass*, *4*, 187–203.

Arnold, J. E. (2013). What should a theory of pronoun interpretation look like? Commentary on Kehler & Rohde (2013): a probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics*, *39*, 59–73.

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

Arnold, J. E., Eisenband, J., Brown-Schmidt, S., & Trueswell, J. (2000). The rapid use of gender information: evidence of the time course of pronoun resolution from eye tracking. *Cognition*, *76*, B13–B26.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

Botting, N., Riches, N., Gaynor, M., & Morgan, G. (2010). Gesture production and comprehension in children with specific language impairment. *British Journal of Developmental Psychology*, *28*, 51–69.

Brooks, P. J., & Tomasello, M. (1999). Young children learn to produce passives with nonce verbs. *Developmental Psychology*, *35*, 29–44.

Capone, N. C., & McGregor, K. K. (2005). The effect of semantic representation on toddlers' word retrieval. *Journal of Speech, Language, and Hearing Research*, *48*, 1468–80.

Chambers, C. G., & Smyth, R. (1998). Structural parallelism and discourse coherence: a test of centering theory. *Journal of Memory and Language*, *39*, 593–608.

Cook, S. W., Duffy, R. G., & Fenn, K. M. (2013). Consolidation and transfer of learning after observing hand gesture. *Child Development*, *84*, 1863–71.

Gernsbacher, M. A., & Hargreaves, D. J. (1988). Accessing sentence participants: the advantage of first mention. *Journal of Memory and Language*, *27*, 699–717.

Gernsbacher, M. A., & Hargreaves, D. J. (1992). The privilege of primacy: experimental data and cognitive explanations. In D. L. Payne (Ed.), *Pragmatics of word order flexibility* (pp. 83–116). Philadelphia, PA: John Benjamins.

Gómez, R. L., Bootzin, R. R., & Nadel, L. (2006). Naps promote abstraction in language-learning infants. *Psychological Science*, *17*, 670–4.

Goodrich, W. (2009). *Gesture as input in language acquisition: Learning 'who she is' from 'where she is'*. (Unpublished doctoral dissertation). University of California, Berkeley.

Goodrich, W., & Hudson Kam, C. L. (2009). Co-speech gesture as input in verb learning. *Developmental Science*, *12*, 81–7.

Goodrich Smith, W., & Hudson Kam, C. L. (2012). Knowing 'who she is' based on 'where she is': the effect of co-speech gesture on pronoun comprehension. *Language and Cognition*, *4*, 75–98.

Goodrich Smith, W., & Hudson Kam, C. L. (2015). Children's use of gesture in ambiguous pronoun interpretation. *Journal of Child Language*, *42*, 591–617.

Gordon, P. C, Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names and the centering of attention in discourse. *Cognitive Science*, *17*, 311–47.

Hartshorne, J. K., Nappa, R., & Snedeker, J. (2011). Ambiguous pronouns processing development: probably not U-shaped. In N. Danis, K. Mesh, & H. Sung (Eds.), *Proceedings of the 35th annual Boston University Conference on Language Development* (pp. 272–82). Somerville, MA: Cascadilla Press.

Hartshorne, J. K., Nappa, R., & Snedeker, J. (2015). Development of the first-mention bias. *Journal of Child Language*, *42*, 423–46.

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin*, *137*, 297–315.

Järvikivi, J., van Gompel, R. P., Hyönä, J., & Bertram, R. (2005). Ambiguous pronoun resolution: contrasting the first-mention and subject-preference accounts. *Psychological Science*, *16*, 260–4.

Karmiloff-Smith, A. (1986). Some fundamental aspects of language development after age five. In P. Fletcher & M. Garman (Eds.), *Language acquisition* (pp. 455–74). Cambridge University Press.

Keenan, E. (1976). Towards a universal definition of 'Subject'. In C. Li (Ed.), *Subject and Topic* (pp. 61–77). New York: Academic Press.

Kehler, A., & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics*, *39*, 1–37.

Kendon, A. (2004). *Gesture: visible action as utterance*. Cambridge University Press.

Kertoy, M. K. (1991). Listening comprehension for sentences: the accessibility of referents for pronouns as a function of age, topic continuity, and pronoun emphasis. *Journal of Experimental Child Psychology*, *52*, 344–53.

Lüke, C., & Ritterfeld, U. (2014). The influence of iconic and arbitrary gestures on novel word learning in children with and without SLI. *Gesture*, *14*, 204–25.

Maquet, P., Peigneux, P., Laureys, S., & Smith, C. (2002). Be caught napping: you're doing more than resting your eyes. *Nature Neuroscience*, *5*, 618–19.

McNeill, D. (1992). *Hand and mind: what gestures reveal about thought*. University of Chicago Press.

Megherbi, H., & Ehrlich, M.-C. (2009). The on-line interpretation of pronouns and repeated names in seven-year-old children. *Current Psychology Letters: Behaviour, Brain & Cognition*, *25*(2). Retrieved from <http://cpl.revues.org/4895>.

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

Morford, M., & Goldin-Meadow, S. (1992). Comprehension and production of gesture in combination with speech in one-word speakers. *Journal of Child Language*, *19*, 559–80.

Mumford, K. H., & Kita, S. (2014). Children use gesture to interpret novel verb meanings. *Child Development*, *85*, 1181–9.

Nappa, R., & Arnold, J. E. (2014). The road to understanding is paved with the speaker's intentions: cues to the speaker's attention and intentions affect pronoun comprehension. *Cognitive Psychology*, *70*, 58–81.

Nelson, K. (1977). Facilitating children's syntax acquisition. *Developmental Psychology*, *13*, 101–7.

Pinker, S., Lebeaux, D. S., & Frost, L. A. (1987). Productivity and constraints in the acquisition of the passive. *Cognition*, *26*, 195–267.

Pyykkönen, P., & Järvikivi, J. (2010). Activation and persistence of implicit causality information in spoken language comprehension. *Experimental Psychology*, *57*, 5–16.

Pyykkönen, P., Matthews, D., & Järvikivi, J. (2010). Three-year-olds are sensitive to semantic prominence during online language comprehension: a visual world study of pronoun resolution. *Language and Cognitive Processes*, *25*, 115–29.

R Core Team (2017). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Online <https://www.R-project.org/>.

Ramscar, M. (2013). Suffixing, prefixing, and the functional order of regularities in meaningful strings. *Psihologija*, *46*, 377–96.

Ryskin, R. A., Fine, A., & Brown-Schmidt, S. (2017). *Do listeners learn speaker/accent-specific syntactic biases?* Poster presented at the 30th Annual CUNY Conference on Human Sentence Processing, Cambridge, MA.

Ryskin, R. A., Qi, Z., Duff, M., & Brown-Schmidt, S. (2017). *Syntactic variability between and within speakers: When to adapt, when to generalize?* Talk presented at the 30th Annual CUNY Conference on Human Sentence Processing, Cambridge, MA.

Sauermann, A., & Gagarina, N. (2017). Grammatical role parallelism influences ambiguous pronoun resolution in German. *Frontiers in Psychology*, *8*. doi:10.3389/fpsyg.2017.01205

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science*, *13*, 13–19.

So, W. C., Coppola, M., Licciardello, V., & Goldin-Meadow, S. (2005). The seeds of spatial grammar in the manual modality. *Cognitive Science*, *29*, 1029–43.

So, W. C., Kita, S., & Goldin-Meadow, S. (2009). Using the hands to identify who does what to whom: gesture and speech go hand-in-hand. *Cognitive Science*, *33*, 115–25.

So, W. C., & Wong, K.-Y. M. (2018). Tracing the development of spatially modulated gestures in the manual modality in nonsigning Chinese-speaking children. *Applied Psycholinguistics*, *39*(3), 527–44.

Song, H., & Fisher, C. (2005). Who's 'she'? Discourse prominence influences preschoolers' comprehension of pronouns. *Journal of Memory and Language*, *52*, 29–57.

Song, H., & Fisher, C. (2007). Discourse prominence effects on 2.5-year-old children's interpretation of pronouns. *Lingua*, *117*, 1959–87.

Stickgold, R., & Walker, M. P. (2005). Memory consolidation and reconsolidation: What is the role of sleep? *Trends in Neurosciences*, *28*, 408–15.

Tomasello, M., Brooks, P. J., & Stern, E. (1998). Learning to produce passive utterances through discourse. *First Language*, *18*, 223–37.

Valian, V., & Casey, L. (2003). Young children's acquisition of wh-questions: the role of structured input. *Journal of Child Language*, *30*, 117–43.

Vogt, S., & Kauschke, C. (2017). Observing iconic gestures enhances word learning in typically developing children and children with specific language impairment. *Journal of Child Language*, *44*, 1458–84.

Wakefield, E., Hall, C., James, J., & Goldin-Meadow, S. (2018). Gesture of generalization: gesture facilitates flexible learning of words for actions on objects. *Developmental Science*, *21*(5). doi:10.1111/desc.12656

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. (2009). Experience and sentence processing: statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*, 250–71.

Yeo, A., Ledesma, I., Nathan, M. J., Alibali, M. W., & Church, R. B. (2017). Teachers' gestures and students' learning: sometimes 'hands off' is better. *Cognitive Research: Principles and Implications*, *2*(41). doi:10.1186/s41235-017-0077-0

Yow, W. Q. (2015). Monolingual and bilingual preschoolers' use of gestures to interpret ambiguous pronouns. *Journal of Child Language*, *42*, 1394–407.

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press

Appendix

Pre- and post-training test

1. Sally and Tim are swimming in a pool. Sally likes to jump off the diving board. Tim doesn't like it when Sally jumps, because she splashes water. (Filler)
   Filler Question: Where are Sally and Tim?
   Filler Question: What does Sally like to do?
   Filler Question: Does Tim like it when Sally jumps?
2. Donald and Mickey are good friends. Donald is walking up a hill, and Mickey is at the top. He has an umbrella, which is good because it looks like it's going to rain. (Experimental item)
   Experimental Question: Who has an umbrella?
   Filler Question: Are Donald and Mickey friends?
   Filler Question: What is the weather like?
3. Goofy and Pluto were walking towards each other on the street. When they saw each other, Goofy smiled, and Pluto grinned. He said, "It's good to see you". (Experimental item)
   Experimental Question: Who said "It's good to see you"?
   Filler Question: What are Goofy and Pluto doing?
   Filler Question: Did they like seeing each other?
4. Bobby gave Andrea a new CD for her birthday. When they listen to the CD, Bobby sings, and Andrea dances. (Filler)
   Filler Question: Who gave Andrea a new CD?
   Filler Question: Does Andrea like the new CD?
   Filler Question: What do they do along with the music?
5. Annie and Sarah are having a picnic in the park. They have a lot of food with them. Annie is carrying the picnic basket, and Sarah has a blanket for them to sit on. She's excited about the cookies. (Experimental item)
   Experimental Question: Who is excited about the cookies?
   Filler Question: What are Annie and Sarah doing?
   Filler Question: Who has a blanket to sit on?
6. Sam and John are good friends, but they live very far apart, and hardly ever get to see each other. Recently, Sam and John had a summer vacation. He went to visit. They were very happy to see each other. (Experimental item)
   Experimental Question: Who went to visit?
   Filler Question: Do Sam and John live far apart?
   Filler Question: Did they like seeing each other?
7. My cat's name is Fluffy. Fluffy has white fur. My neighbor's cat is named Snowball. Snowball has white fur. Fluffy and Snowball don't like each other, he's afraid of him. (Experimental item)
   Experimental Question: Who is afraid of the other cat?
   Filler Question: What is my cat named?
   Filler Question: What color is Snowball?
8. It's a warm sunny day. Billy and Sam are playing catch with a baseball. Billy isn't very good at catch because he drops the ball a lot. (Filler)
   Filler Question: What is the weather like?
   Filler Question: Who is playing catch?
   Filler Question: Is Billy good at playing catch?

Cite this article: Goodrich Smith W, Black AK, Hudson Kam CL (2019). Learning speech-internal cues to pronoun interpretation from co-speech gesture: a training study. *Journal of Child Language* **46**, 433–458. https://doi.org/10.1017/S0305000918000557

https://doi.org/10.1017/S0305000918000557 Published online by Cambridge University Press