

Who did what to whom: Developmental perspectives on the meaning and communication of transitive events

Melissa Kline

Sc.B., Brown University (2008)

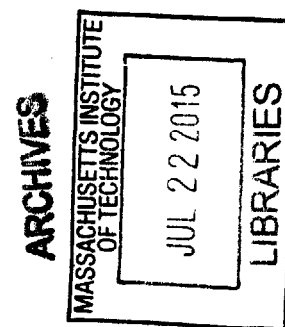
Submitted to the Department of Brain and Cognitive Sciences in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

at the MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2015

© Massachusetts Institute of Technology 2015. All rights reserved.



Signature redacted

Signature of Author: \_\_\_\_\_

Department of Brain and Cognitive Science

June 5, 2015

Certified by: \_\_\_\_\_  
Signature redacted

Edward Gibson

Professor of Cognitive Science

Thesis Supervisor

Accepted by: \_\_\_\_\_  
Signature redacted

Matthew Wilson

Sherman Fairchild Professor of Neuroscience

Director of Graduate Education for Brain and Cognitive Science

Who did what to whom: Developmental perspectives on the meaning and communication of transitive events

by

Melissa Kline

Submitted to the Department of Brain and Cognitive Science

on March 23, 2015, in partial fulfillment of the

requirement for the degree of

Doctor of Philosophy

## ABSTRACT

Human language is notable for its expressivity; syntax is powerful and allows for potentially unlimited new sentences. But even simple transitive sentences like “I broke the lamp” provide a sophisticated tool for communication, capture the basic building blocks of syntax and semantics that are widely agreed to be part of our linguistic capacity like *agent* or *subject*. With this relatively simple machinery, we are able to move a cognitive representation of an event from one person’s head to another. How is this possible? In this dissertation, I examine both adult and child language to understand this capacity. Paper 1 examines the link between non-linguistic cognition and preschoolers’ expectations about the meaning of novel verbs. We find that even though transitive verbs can refer to many event types, 3- and 4-year-olds are more likely to associate them with scenes with spatiotemporal features indicating causation. Papers 2 and 3 ask a second question: how do people organize language to facilitate communication? Paper 2 probes how adults order the basic elements (Subject, Verb, Object) in a task that appears to be independent of native language constraints, and tests whether the *content* of the message leads gesturers to reorganize their utterances. Paper 3 asks whether adults and children are aware that the relative informativity of arguments depends on context, and whether they can successfully make decisions in a novel communication task. By limiting the expression of transitive sentences to just two words (e.g. MONKEY EAT), we discover which elements people consider to be most informative. Both adults and children flexibly adjust their expectations about informative sentences according to which arguments are the most ambiguous in context. Together, these case studies help us understand how human language accomplishes

its communicative goals, both in terms of the cognitive representations recruited for processing complex events in language, and the strategies used for expressing them. Whatever the formal nature of the representations involved in syntax and semantics, they must ultimately allow us to form predicates over nonlinguistic representations of the world, and they must support the kinds of pragmatic inferences that we know people can make.

Thesis Supervisor: Edward Gibson

Title: Professor of Cognitive Science

## ACKNOWLEDGEMENTS

I am incredibly grateful for the hard work, energy, and dedication that my mentors have put into both the scientific work that appears in this dissertation and my development as a researcher over the past six years. I have been fortunate to work with faculty who have encouraged my independence from the very start. My co-advisors Ted Gibson and Laura Schulz have been my anchors through this process, helping me to ground my work deeply in multiple research communities and to think like a cognitive scientist. Running up and down the 3<sup>rd</sup>-4<sup>th</sup> floor stairs between their labs has been an exercise in juggling projects, work styles, and collaborators, and I feel very lucky to already have had this broad base of experience in how labs and research communities work. Their flexibility and openness in making this possible has played a key role in how I've developed my research program, as has their shared insistence on clarity of explanation, every single meeting, every single time. They also both throw great lab parties, though we only get to watch *The Bachelor* at Ted's house. Josh Tenenbaum has always been available for those infamous 'quick chats' that leave me with new research directions and months of mulling to do, and his insights and sheer level of interest in cognitive science have been inspiring from the very beginning. Finally, Jesse Snedeker at Harvard has been a wonderfully generous mentor and collaborator – she manages to be one of my most rigorous critics and most supportive voices at the same time, thanks to her openhearted, genuine interest in the projects I've brought to her and her unfailing willingness to see where the data takes us, wherever that might be.

Grad school has brought me into contact with wonderful, intelligent people who have been ready throughout my time here to offer contributions & insights, traipse around conferences, and star in stimulus videos. They include: Paul Muentener, Ev Fedorenko, Steve Piantadosi, Celeste Kidd, Leon

Bergen, Kyle Mahowald, Richard Futrell, Tim O'Donnell, Idan Blank, Hyo Gweon, Mike Frank, Kim Scott, Rachel Magid, Julian Jara-Ettinger, Pedro Tsividis, Yang Wu, Julia Leonard, Jorie Koster-Hale, Lindsey Powell, Hilary Richardson, Amy Skerry, Sammy Floyd, Danny Friel, Ali Horowitz, Eva Wittenberg, Annemarie Kocab, Tanya Levari, Roman Feiman, Amy Geojo, Josh Hartshorne, & Manizeh Khan. All of these people have made my time at MIT wonderful so far, and I feel very lucky to not be leaving them entirely quite yet. This dissertation would not have been possible without NSF and MIT fellowship support, and it especially would not be possible without the children and parents at the Boston Children's Museum, and participants through Mechanical Turk and the MIT study pool who contributed their time to make this work possible. The MIT UROP program has also introduced me to a series of wonderful undergraduates from both MIT and elsewhere who have poured their hours and their brains into these projects, many of them for multiple semesters. They truly made this work possible - especially the developmental components - and in many cases their insights contributed significantly to the final shape of projects. I'd like to especially highlight Emily Kasman, Audra Podany, Nora Darago, Cristina Spicer, Miguel Salinas, Jenny Sangliana, and Chelsea Lide for their extended participation and thoughtful contributions at various points over the past six years.

Of course, many of the people who got me to this place I met long before I came to MIT. At Brown I was lucky to work with Katherine Demuth, who encouraged my interests and was willing to let a sophomore dive right into research right away and gave me my first tastes of the scientific process. Before that, my time at Northwest in Seattle gave me the best foundation I could have asked for - my teachers there, with special thanks to Mark Terry and Renee Fredrickson, created a wonderful learning community and took me and my questions seriously. Thanks to them I know that 20 minutes is not a distance and the importance of courtesy and common sense. Lastly, my family is filled with curious people who are always happy to get into hours-long discussions about airplane wings or particle physics - my parents were my first science teachers and continue to be my cheerleaders from all the way across the country.

Somerville has been a wonderful place to live over the last six years, and has included a wide circle of friends outside of science (SCIENCE) who have made my time here an absolute joy, even as some of them have traveled on to other places. They have kept me sane, they have taken me to Iceland, they have drunk my dubious homebrews and provided me with same, played more boardgames than I can count, and are all around wonderful neighbors, near & far. Again, I'm thrilled I'll get to keep hanging around them.

Finally and most of all, this thoosis is dedicated to my husband, Ben Kline Struhl. <4 <4 <4.

## TABLE OF CONTENTS

<b>Abstract</b> .....	<b>2</b>
<b>Acknowledgements</b> .....	<b>4</b>
<b>Chapter 1 - Introduction</b> .....	<b>8</b>
<b>Chapter 2 – Linking Language and Events: Spatiotemporal cues drive children’s expectations about the meanings of novel transitive verbs</b> .....	<b>13</b>
<b>Abstract</b> .....	<b>13</b>
<b>Introduction</b> .....	<b>14</b>
<b>Experiment 1 – Spatiotemporal Continuity</b> .....	<b>19</b>
<b>Experiment 2 – Temporal Order</b> .....	<b>26</b>
<b>General Discussion</b> .....	<b>33</b>
<b>Chapter 3 – Do word order patterns in gesture reflect general communicative patterns or modality-specific production constraints?</b> .....	<b>39</b>
<b>Abstract</b> .....	<b>39</b>
<b>Introduction</b> .....	<b>40</b>
<b>Experiment 1: Comprehension</b> .....	<b>49</b>
<b>Experiment 2: Production with Case marking</b> .....	<b>56</b>
<b>Experiment 3: Production without Body-Based Gestures</b> .....	<b>72</b>
<b>General Discussion</b> .....	<b>77</b>

**Chapter 4 - Adults and 5-6-year-olds make information-sensitive choices about argument dropping in transitive sentences ..... 82**

**Introduction ..... 82**

**Experiment 1 - Adult production with '2 vs. 1' contrast sets ..... 87**

**Experiment 2 - Adult production with 'many vs. many' contrast sets..... 92**

**Experiment 3 - Child comprehension/speaker choice with '2 vs. 1' contrast sets ..... 97**

**General Discussion.....102**

**Final Discussion..... 106**

**Appendix A - Online Repositories for Data and Stimuli ..... 111**

**References ..... 112**

## CHAPTER 1 - INTRODUCTION

Human cognition is notable not just because we can think new thoughts about complex concepts and abstract relationships, but because we can get those thoughts into someone else's head. Language has long been recognized as a key and possibly unique ability of the human race because of its complex formal structure, but its critical function is to facilitate this communication of concepts. How do babies grow into cognitive creatures who can accomplish this? While animal communication allows for the representation of single objects or locations (cf. Scott-Phillips, 2015), children learning language go far beyond this in the first three years. Even very simple sentences like

*The Ewok broke the droid*

require syntactic and semantic **structure**, that is, predictable relationships between words that allow us to express thoughts not just about objects in the world, but about how they interact with one another and change over time. To construct or comprehend these sentences, children must learn to understand these structures, knowing enough about syntax, semantics, and the predictable links between them to combine the constituent words in the right way. Perhaps less obviously, they must have some idea about what it is they are trying to communicate – they need the mental capacity to think about the event and form a proposition about it, which will require both specific knowledge of particular concepts (like *breaking*) and more general knowledge of the types of relationships (like X CAUSE Y) that can be expressed in language. Finally, communicating successfully requires decision-making about what to say and when, such that a conversational partner in a rich social setting will draw the appropriate inferences about what you are talking about.



In this dissertation, I take a wide developmental approach to discovering how these challenges are accomplished, examining both adult and child language use to understand the basic processes that connect language to the rest of cognition throughout the lifespan. I develop these ideas through a common case study, the simple transitive sentence (1). This simple sentence type, though lacking many of the features that have been of interest to linguists such as clause hierarchies or constituent movement, captures the use of the basic building blocks of syntax and semantics, representations that are widely agreed to be part of our linguistic capacity such as the concept of an *agent* or *subject*. Thus, they represent one of the simplest cases in language where multiple linguistic representations must be coordinated.

In particular, while the transitive in English can take on a wide variety of meanings, I focus on a particular class of events which is both stably used in the transitive across language and related closely to event paradigms in core cognition: that of a human agent acting on an object in order to bring about some state of the world (e.g. obtaining the object, changing it, etc.) In non-linguistic cognition, such scenes are interesting to children from very early in infancy, and they learn to reason flexibly about them during the first two years of life (for instance making predictions from a person's behavior back to their goal, or from their preferences to a prediction about likely actions, cf. Csibra, 2003; Woodward, 1998).

Using this case study, then, I attempt to understand how human language accomplishes its communicative goals, both in terms of the cognitive representations that are recruited for processing complex events in language, and the strategies that speakers and listeners use for expressing them. I do this by asking questions about how language users of different ages deal with these simple two-place predicates when they are contrasted with other sentences (e.g. *Vader fell*) or other possible events (e.g. a choice between agents – *Han/Leia/Luke broke the droid*).

Three projects are included in this dissertation. In Paper 1, I ask what expectations preschoolers have about the relationship between transitive sentences (syntax) and causality, an aspect of meaning which has been critical for theories of semantics but which has not been tied closely to nonlinguistic theories of cause. The aim of this study is to understand how lexical semantics – the aspects of verb meaning which are closely tied to the syntactic structure of sentences – relates to the areas of early-developing cognition that overlap closely in conceptual content. Understanding the relationship between semantics and combinatorial syntactic rules has been a central goal of linguistic theories in lexical semantics and syntax since the 1960s (c.f. Dowty, 1991; Fillmore, 1968; Goldberg, 2006; Hale & Keyser, 1993; Jackendoff, 1983; Levin & Rappaport Hovav, 2005; Talmy, 1985). What is key for the present work is the fact that these systematic regularities do exist, and they could have been otherwise than they are. The aspects of

verb meaning that affect syntactic structure do not include conceptual dimensions such as color which are nevertheless available to young children. Understanding when and how the existing regularities form is therefore a key goal for understanding language acquisition and its connection to the rest of cognitive development.

I focus specifically on causation, an aspect of meaning which is both central for interpreting transitive sentences and available to young children in at least a basic form. In two studies, we present 3- and 4-year-old children with a novel sentence (e.g. *the girl is wugging the toy*) paired with two possible referents: one which presents a simple causal event, and one which maintains the participants and their actions, but changes the spatiotemporal relationship between the sub-events. While we know that children distinguish among broad semantic classes for which the transitive can and cannot be used, we know very little about the cognitive representations that underlie these preferences. If young children's representations of the possible meanings for transitive sentences make reference to their nonlinguistic understanding of CAUSE, we expect them to selectively prefer causal scenes when they hear positive transitive sentences, but not when they hear control sentences. On the other hand, if children's expectations about the meanings of transitive sentences are not tied to causation (tapping instead either a broader category of two-participant events or a notion of causation that is distinct from the nonlinguistic one), we do not expect to see a specific preference for the causal scenes.

In Paper 2, I ask how adults choose to order the basic components of a transitive sentence – agent, verb, and patient – in a task that may free them from the constraints of their native language. While word order varies across the world's languages, it does so unevenly: over 80% of existing languages are either SVO or SOV order, a fact that cannot be accounted for merely by historical factors (Dryer, 2005). Previous work from our lab and others has suggested that these order may predominate either because they are cognitive natural, or because they offer an advantage for communication (Gibson, Piantadosi, et al., 2013; MacWhinney, 1977). In essence, this work attempts to find connections between the basic argument structure of a sentence and principles of rational communication which have been found guide other aspects of language such as reference choice and sentence interpretation .

The goal of Paper 2 is thus to understand how the basic word order of a transitive sentence relates to the communicative goal of expressing its argument structure successfully. In a gesture-communication paradigm, three experiments examine whether the basic word order of transitive events is sensitive to noisy-channel communication constraints. In particular, they examine the hypothesis that SVO order arises because it is better suited for successful communication (i.e. it decreases the likelihood that the

identity of the agent and patient will be mistaken), particularly in the case when both agent and patient are animates. We test whether this animacy-dependent shift to SVO (previously established) responds to changes in the task that alter the communicative affordances of the gestures, by providing participants with a case system that can be used to mark the identity of the subject and object independent of word order. We predicted that such a change, which lowers the possibility of confusion of thematic roles, would also lower the incidence of SVO order. In the course of these experiments we discovered a series of possible confounds arising from production constraints on gesture; the later experiments and analyses therefore focuses on what we can and cannot do to disentangle the effects of native language, production constraints, and pragmatic pressures on basic word order.

Paper 3 also focuses on the intersection of argument structure and communication, asking how adults and children take possible alternative events into account when deciding what to say. One of the key differences between noun and verb learning is that verbs are usually transient in time, so that the verb is often not available in the context for the child when it is spoken (cf. Gleitman, 1990; Pinker, 1989). Between this and the fact that understanding verb reference almost always requires identifying event participants as well, tailoring these sentences for the needs of the listener may be especially important. When faced with grammatically appropriate alternatives and referential communication tasks, adult speakers select words and constructions to make utterances easy to understand (Frank & Goodman, 2012; Jaeger, 2010; Piantadosi, Tily, & Gibson, 2011). However, the research on referential expressions has focused primarily on ways of referring to objects, rather than events. In this work, I instead ask whether adults and children are sensitive to both the status of different entities as potential agents and patients, and to the relative information value of event participants in context.

To do this, I ask adults and young children to work with phrases that describe transitive events like *The monkey eats the orange* but include only two words (e.g. MONKEY EAT). This necessitates dropping one of the central content words; we focus in particular on cases where it is one of the event participants (agent or patient) that gets dropped. Using this paradigm, we then evaluate attempts to communicate when a speaker must consider not just a single event, but an array of possible events. In particular, we ask how adults and children make decisions when a sentence must detect the correct participants out of a set of possible agents and possible patients. Depending on the context, a phrase like MONKEY EAT could be more or less informative: if the monkey is surrounded by several different kinds of fruit, it is not very informative; saying “she’s eating the orange” would be very informative. But if the monkey, the duck, and the donkey are all standing around an orange, “she’s eating the orange” is not very informative at all.

In two experiments with adults, I ask what people choose to say when they are faced with using shortened transitive sentences with different arrays of agents and patients. In an experiment with 5-6 year olds, I then ask whether children know which of two puppet speakers is the better ‘helper’ when both give only a partial description of a scene. This line of work will bring our understanding of how basic argument structure is processed in language in line with modern information theoretic perspectives, and produce paradigms that will allow for much more specific analyses and modeling of early child language.

These three papers each focus on a different aspect of the problem of successfully communicating a canonical agent-acts-on-patient event in the form of a transitive sentence. Together, they provide some new perspectives on the links between language and the basic cognitive capacities underlie its expressive power. These links are of two types, roughly: the links between sentence structure and the meanings they express, and the links between those same structures and the general capacities that humans have for understanding each others’ minds and communicating successfully. Whatever the formal nature of the representations involved in language-specific syntax and semantics, they must ultimately allow us to form predicates over nonlinguistic representations of the world, and they must support the kinds of pragmatic inferences that we know people can make. By understanding both of these points of integration, we will gain a fuller understanding of how we manage, even from young childhood, to get a perspective on an event from one mind to another.

## CHAPTER 2 – LINKING LANGUAGE AND EVENTS: SPATIOTEMPORAL CUES DRIVE CHILDREN’S EXPECTATIONS ABOUT THE MEANINGS OF NOVEL TRANSITIVE VERBS

### ABSTRACT

How do children map linguistic representations onto the conceptual structures that they encode? In the present studies, we provided 3-4 year old children with minimal-pair contrasts in order to determine the effect of particular event properties on novel verb learning. Specifically, we tested whether spatiotemporal cues to causation also inform children’s interpretation of transitive verbs. Unlike previous studies of this type, we manipulated specific scene cues, rather than contrasting event category prototypes. In Experiment 1, we examined spatiotemporal continuity. Children saw scenes with puppets that approached a toy in a distinctive manner, and toys that lit up or played a sound. In the causal events, the puppet contacted the object, and activation was immediate. In the noncausal events, the puppet stopped short before reaching the object, and the effect occurred after a short pause (apparently spontaneously). Children expected novel transitive verbs to refer to spatiotemporally intact causal interactions rather than to 'gap' control scenes. In Experiment 2, we manipulated only the temporal order of sub-events, holding spatial relationships constant. Children mapped transitive verbs to scenes where the agent's action closely preceded the activation of the toy over scenes in which the timing of the two events was switched. These studies reveal that children’s expectations about transitive verbs map closely to their nonlinguistic understanding of causal events: children expect transitive syntax to refer to scenes where the agent's action is a plausible cause of the outcome. These findings open a wide avenue for exploration into the relationship between children’s linguistic and non-linguistic knowledge.

## INTRODUCTION

During their first several years of life, children develop rich and robust cognitive models of the world around them. They represent the differences between people and objects, make predictions about physical and social events, and learn complex patterns of causal information (C. Baker, Saxe, & Tenenbaum, 2011; Luo, Kaufman, & Baillargeon, 2009; Sobel, Tenenbaum, & Gopnik, 2004; Spelke, 1990; Woodward, 1998). At the same time, they learn the structure and content of at least one language, acquiring representations that allow them to understand the speech around them and produce novel sentences of their own. Relatively little is known about how children map these linguistic representations onto the conceptual structures that they encode.

A central property of human language is that there are systematic correspondences between the form of sentence (syntax) and its meaning (semantics) (M. Baker, 1988; Roger Brown, 1958; Fillmore, 1968; Montague, 1970; Pinker, 1984). For instance, although there are no familiar content words in the sentence *The blicket daxed the wug to the boof*, adults can infer that the event involved transfer and that *dax* means something kind of like *give* (Gilette, Gleitman, Gleitman, & Lederer, 1999; Kako, 2006; Snedeker & Gleitman, 2004). Linguistic studies, within and across languages, have revealed a rich set of connections between semantics and syntax that are captured in the *argument structures* of verbs (cf. Levin & Rappaport Hovav, 2005 for a review). These correspondences raise several questions about how event representations and linguistic forms interact during development. What do young children learn about events from the syntax of sentences describing them? And critically, do children's expectations about the meanings of verbs follow or depart from their intuitions about non-linguistic concepts? In the present studies, we provide children with minimal-pair event contrasts in order to explore a specific feature, causation, which may guide young children's expectations about transitive sentences like *Sarah broke the lamp*. Studies of causal perception have found that the spatiotemporal relationships between events are a critical cue to causation. Here, we examine whether these same spatiotemporal cues also inform children's interpretation of transitive verbs.

Understanding how children map between causal concepts and language is critical for two reasons. First, causal concepts underlie some of the central generalizations about the form and interpretation of language. Causal information is carried not only by particular words in a language (such as *make* or *because*) but also by the form of a sentence. A sequence like "*Ben pilked the cup. The cup pilked.*" suggests to adult participants that *pilking* involves changing or causally affecting the cup in some way,

even through the actual verb is unfamiliar (Kako, 2006). In lexical semantics, this fact is captured by theories of verb meaning which encode causation as a representational primitive, as in:

- (1) a. The box opened
- b. [BECOME [Y open]]
  
- (2) a. John opened the box
- b. [X CAUSE [BECOME [Y open]]] (cf. Jackendoff, 1983)

While there are a variety of proposals about lexical semantics, which differ in many respects, most of these theories break the meanings of verbs into pieces (subpredicates) and include a subpredicate that encodes cause (Croft, 2012; Fillmore, 1968; Folli & Harley, 2008; Hale & Keyser, 1993; Jackendoff, 1990; Pinker, 1989; Rappaport Hovav & Levin, 1998; Talmy, 1988; Van Valin & LaPolla, 1997).

Argument structure provides a critical set of constraining hypotheses for a person learning a new verb. As shown above, semantic structures dictate the number of arguments expressed about an event and the hierarchical relationships between them, and these relationships are reflected in the syntax of sentences. Thus, when *running*, *chasing*, *playing*, and *laughing* are all going on in a scene, the sentence surrounding a new verb can provide important information about the number of participants and the nature of their relation, that can allow the child determine which specific event is being referred to.

The second motivation for understanding how causation maps to language is that children acquire much of their causal knowledge about the world through language. Testimony – second-hand information provided by adults or peers, the term used in causal development research for such second-hand causal information (Harris, 2002), can change children’s understanding of causal events in a variety of ways. Children shift from perceptually-based to causally-based categorization if causal language is used to describe objects (Nazzi & Gopnik, 2000, 2001), preschoolers explore perceptually identical objects with disparate causal properties more if the objects are given the same name (Schulz, Standing, & Bonawitz, 2008), and 2-year olds who have learned that one event *predicts* another only try themselves to use the first event to cause the second if the relationship has been described with causal language like ‘the block *makes* the helicopter spin’ (Bonawitz et al., 2010).

Despite the fact that children learn a great deal about the world around them from this second-hand testimony (Harris & Koenig, 2006), relatively little is known about how children map their nonlinguistic causal understanding onto language. Children’s early causal understanding has been analyzed in a number

of frameworks, including causal Bayes nets that track and explain patterns of covariation between events, and force-dynamic theories which focus on directed physical interactions between objects (Göksun, George, Hirsh-Pasek, & Golinkoff, 2013; Gopnik et al., 2004; Pearl, 1988; Wolff & Song, 2003). There is rich body evidence showing that, for infants and adults, spatial and temporal cues play a critical role in distinguishing causal events from noncausal events (Leslie, 1984; Leslie & Keeble, 1987; Michotte, 1963; Muentener & Carey, 2010). Infants' causal perception has primarily been studied with simple motion events (e.g., one billiard ball rolling into another and causing it to move). Critically, the perception of causality depends on the spatiotemporal relationship between the two sub-events (in this case, the first ball's motion and the second ball's motion). If the second ball pauses before moving away, or if the first ball does not make contact with the second before it starts moving, adults do not perceive a causal relationship (Michotte, 1963). Infant causal perception appears to be guided by the same constraints. For example, when 6 month olds see an intact causal event, they dishabituate strongly to a new event where the first ball (the causal agent) changes (Leslie & Keeble, 1987). In contrast, if they see an event with a spatiotemporal gap, they are less impressed by changes in the first ball. The fact that very young children use this kind of fine-grained spatiotemporal information to understand causal events indicates that they are sensitive to the *sub-event structure* of these events. They do not group all two-participant 'billiard-ball' scenes together or treat each sub-event as a separate entity. Instead they recognize that the relationship between the motion of the first and the motion of the second ball carries crucial information about the event as a whole.

Transitive sentences like *Sarah broke the lamp* are a critical test case for understanding how children might use information about argument structure to learn verb meanings. Causal events (as described in sentences like 2) are just one kind of two-participant event. Thus the number of arguments in the sentence does not provide strong evidence that sentence describes a causal event --just like the number of billiard balls in a display does not provide strong evidence that the event itself is causal. However, both within and across languages, causal events are more likely than noncausal events to be described in a transitive sentence (3) rather than a sentence like (4), where intransitive verb appears with a prepositional phrase.

(3) John broke the vase.

(4) John talked to Bill.

The connection between transitivity and cause is fairly weak in English; transitives can also be used to describe contact (*the lamp touches the table*), perception (*the girl sees the lion*), and spatial relation (*the*



*wall surrounds the castle.*), all of which lack the cause predicate (cf. Levin & Rappaport Hovav, 2005). When we look across languages, however, the picture is clearer: events of direct external causation are consistently described with transitives, whereas the encoding of noncausal events is more variable. For instance, the Russian translations of sentences like ‘The supervisor *manages* the department’ are not transitive sentences but instead have oblique arguments (‘The supervisor *manages over* the department’; Levin & Rappaport Hovav, 2005). This connection between cause and transitivity appears to influence the behavior of adult English speakers, despite the looseness with which English uses transitive syntax. When adult speakers of English are asked to make guesses about the meanings of novel transitive verbs, they interpret them as if they were causal, inferring that the subject exerted force and caused something to move or change, while the object underwent some change of state (Kako, 2006).

The polysemy of English transitive syntax (i.e. the fact that the same surface syntax allows both causal verbs like *break* and noncausal verbs like *touch*) presents an important learning challenge for young children. Children will hear transitive verbs with a variety of meanings: some verbs will encode cause and effect, but others will encode contact, perception, or possession. Thus the evidence that children receive is consistent with two hypotheses. 1) The transitive frame has multiple meanings, each of which encodes a different event structure. On this hypothesis, we might expect the specific mapping between transitive syntax and causation to be acquired earlier than many of the others, both because it relies on concepts that are available early in infancy (see above) and because it is cross-linguistically robust (raising possibility that it serves central role in organization of argument structure). 2) The transitive frame could have a single broad meaning. For example, transitive syntax could be seen as applicable to any event that involves two participant roles. This single mapping would cover not only causal transitives but also transitives encoding contact, perception and spatial relation. In fact, Fisher and colleagues have suggested that infants begin verb learning with precisely this kind of broad mapping strategy (Fisher, 1996; Fisher, Gertner, Scott, & Yuan, 2010; Lidz, Gleitman, & Gleitman, 2003; Yuan, Fisher, & Snedeker, 2012).

The existing research on verb learning in young children is compatible with both of these possibilities. Three findings are particularly relevant. First, children interpret transitive verbs as referring to prototypical causal events (e.g., one girl spinning another girl on a chair) rather than events with parallel action (e.g., two girls jogging) or a single participant (Arunachalam, Escovar, Hansen, & Waxman, 2013; Arunachalam & Waxman, 2010; Naigles, 1990; Yuan & Fisher, 2009; Yuan et al., 2012). This finding has been extensively replicated across a variety of ages and discourse conditions (c.f. Fisher et al., 2010 for a review). Second, children also interpret transitive verbs as referring to prototypical contact events

(e.g., a girl patting another girl on the head) rather than events with parallel actions (Naigles & Kako, 1993). Finally, under some circumstances, children prefer to interpret transitive verbs as referring prototypical causal events rather than prototypical contact events (Naigles, 1996; Scott & Fisher, 2009).

This research is consistent with both the possibility that children have narrow, specific semantic mapping and the possibility that they have a single broad mapping. On the narrow hypothesis, it suggests that they have mapped the transitive construction to both cause and contact (but may prefer the causal mapping). On the broad ‘two participant’ hypothesis, these findings would indicate that both prototypical causal and contact events are seen as having two distinct roles, but that the parallel-action scenes are not. Under this hypothesis, the participant roles in a causal event might be analyzed as somewhat more asymmetric (i.e. having more distinguishing features), and thus causal events are a better candidate for transitive verb meanings.

Critically, while the studies described above have used prototypically causal events, they do not demonstrate that it is the causal properties of an event *per se* that lead to the transitive mapping. These studies were largely designed to explore verb learning and the nature of early syntactic representations, and consequently the stimuli were not designed to target specific conceptual distinctions. This method of pitting two event prototypes against each other originates with Naigles (1990). In this study, children saw two test scenes: one canonically causal scene in which a duck pushing on a bunny’s shoulders to make it bend over, and one ‘parallel action’ scene in which the duck and bunny simultaneously and separately wave their arms. When 2-year-old children had heard sentences like “The duck pilks the bunny” they looked preferentially at the canonically causal scene rather than the parallel-action one. This looking-time preference disappeared when participants heard sentences like “The duck is pilking” instead. What about these events guides toddler’s syntax-specific expectations? Children might be tuning in to many different features of the scene, such as a particular forceful action, the timing of the duck’s shoulder-pushing and the bunny’s bending, or the simple fact that the actions performed by the duck and bunny differ. In other words, causation is conflated both with a particular kind of sub-event (i.e. shoulder-pushing), and with a particular relationship between sub-events (i.e. the difference, characteristic timing, and/or physical proximity of the duck and bunny’s sub-events.)

Thus, the existing studies of children’s early syntactic awareness do not tell us whether children selectively expect transitive verbs to refer to causal events, *qua* causal events. While we know that children expect two participants to be ‘involved’ in an event described with a transitive verb, the nature of this involvement is unclear (Yuan et al., 2012). Do children have a specific mapping between transitive

sentences and causal scenes? To explore this question, we investigated whether 3- and 4-year-old children's expectations about novel transitive sentences are specifically sensitive to the spatiotemporal cues to causality. To do this, we provide children with minimal-pair contrasts, to determine the effect of specific event properties on novel verb learning. We test these questions with 3- and 4-year-olds, who have rich representations of causality in nonlinguistic domains and who are actively building their verb lexicons. In Experiment 1, we examine spatiotemporal *continuity*. All stimuli consisted of a puppet who approached a toy in a distinctive manner (sub-event A), and a toy that lit up or played a sound (sub-event B). In the causal versions of each event, the puppet contacted the object, and activation was immediate. The noncausal event versions were identical except that the puppet stopped short before reaching the object, and the effect occurred after a short pause (apparently spontaneously). Children viewed both events, heard a sentence with a novel verb, and had to choose one of the two scenes. In Experiment 2, we examine temporal *order*. In these stimuli, an actor approached or gestured at an object (sub-event A); the object immediately lit up or moved (sub-event B.) In the noncausal version of each stimulus, the events were identical except that the actor initiated her movement *after* the toy activated.

With the exception of these changes (which are critical to causal perception), the scenes used in both experiments are identical. Thus they were equated for levels of activity, participant asymmetry, and any other factors which might reasonably identify a scene as member of a broad 'two participant' category. The two hypotheses about children's initial expectations about transitive-event mappings thus make very different predictions. If children map transitive sentences specifically to causal scenes (perhaps alongside other specific, possibly weaker mappings) then after hearing novel transitive verbs, they should choose the event with spatiotemporal cues consistent with causal relationships. Under the theory that children make a single, broad mapping between transitive sentences and two-participant scenes, we might expect children to show no preferences between these two scenes.

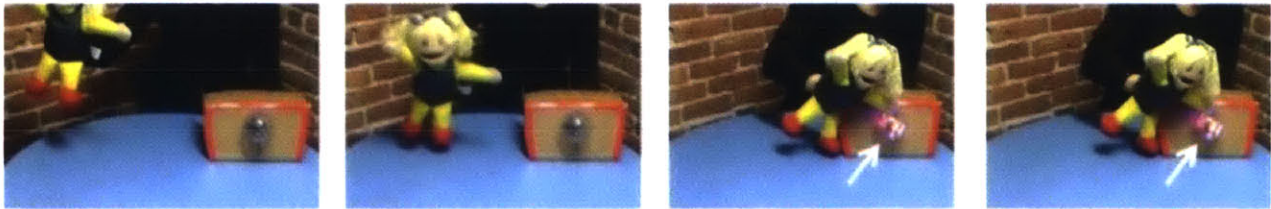
## EXPERIMENT 1 – SPATIOTEMPORAL CONTINUITY

Experiment 1 is the first study to investigate how children's expectations about the meaning of a transitive verb are affected when only a single aspect of the event structure, spatiotemporal contiguity, is varied. All events in this study consisted of scenes containing two sub-events: an action performed by an animate entity (a female puppet), and an outcome effect in a novel toy (e.g. lighting up, spinning around.) Children were presented with minimal pair contrasts, identical to each other save for the spatiotemporal relationship between the sub-events. In the 'no-gap' scenes (see Figure 1), the puppet approached and made contact with the toy, and the effect took place immediately. In the 'gap' contrasts, the puppet

approached the toy but stopped several inches away from it; after a short pause the toy activated, apparently spontaneously. To control for the possibility that children might select the causal events simply because they might be more salient than the noncausal events, all children were asked to identify scenes in which the puppet *wugged the round thing* and *didn't wug the round thing*, with questions presented in random order.

Previous research has already established that children use spatial contact and contingent timing between an action and an outcome to detect causal and noncausal events beginning in infancy. What is at issue here is whether this distinction is relevant to how children interpret the meaning of transitive constructions. If children expect transitive sentences to refer to causal scenes in particular, then when they hear transitive sentences like *Sarah wugged the round thing*, they should choose the (causal) no-gap scenes over the (noncausal) gap variants. On the other hand, if children are sensitive only to coarser scene features such as the active presence of two participants, then children might choose between the events randomly. In particular, since the content of the two sub-events is identical between event versions, children cannot depend on particular features of the entities or the sub-events (e.g. a particular intentional

### “No-Gap” (Causal) Scene



-----

### “Gap” (Non-Causal) Scene



Figure 1: Schematic of the event contrasts used in Experiment 1. Each novel event was filmed in a causal version and a “gap” control introducing a spatial gap and short pause between the agent’s action and the toy’s outcome.

motion by the puppet, or a particular kind of toy action) to guide verb preferences. Any preferences that children show in this task must therefore be due to the differing *relationship* between the sub-events.

## **METHOD**

**Participants:** Preschoolers were recruited from a local children’s museum (n=24 mean age: 3;11, range 3;0-4;9, 12 girls). Participants were replaced if they were unable to reach criteria on the pretest training (n=3.) Five additional children were replaced due to refusal to participate or parental interference. All children received a sticker and award certificate for their participation at the end of the session.

**Materials:** Gap and no-gap videos were created for six novel events. All events were initiated by a female puppet held by the experimenter. In each no-gap event, the puppet contacted a novel apparatus that immediately moved, lit up, or made noise. The ‘gap’ version of each event differed only in the spatiotemporal relationship between action and outcome sub-events, with a roughly 10-15 cm gap and 1-second pause between the puppet’s final position and the activating toy.

In all videos, the event was played through three times, ending on a still shot showing the result and the final position of the puppet. Videos varied between 4.5 and 8 seconds in total length, with no more than a 1 second length difference between the causal and noncausal version of the same event. Descriptions of the events are shown in Table 1, and videos of all stimuli are available at <http://web.mit.edu/mekline/www/Spatiotemporal/>.

Video stimuli were presented on a 17-inch laptop, using the Psychtoolbox extensions of Matlab (Brainard, 1997). An additional apparatus was used during the introduction, consisting of an open-backed box with a toy helicopter on top that could be covertly activated.

**Procedure:** Each session consisted of an introduction, a pretest and the main novel verb test. During the introduction, children were introduced to ‘my friend Sarah’, a puppet who liked to say silly words. The experimenter showed them the helicopter apparatus, demonstrating that *Sometimes, Sarah puts her hand here [on top of the box] and makes it go...But sometimes, it just happens on its own, because there’s a battery inside*. Children were then prompted to activate the toy, and shown again that it could activate spontaneously. Then the experimenter prepared children for the rest of the session by explaining that in the movies they would see, *Sometimes, Sarah makes something happen. Like this, when they’re touching. And sometimes they don’t touch and it just happens on its own, because there’s a battery inside*. Note that

children never heard causal (or noncausal) events described with transitive sentences such as *Sarah's touching the box*.

Following this introduction, children moved to the video presentation, beginning with two pretest trials using the first two events (see Table 1). The pretest was designed to train children on the task and determine whether they understood the events they were seeing. During the pretest trials, children viewed both the gap and no-gap version of one of the novel events, with version and side presentation counterbalanced between children. During these videos, children heard neutral language directing their attention to the video (*Look over here! Whoa, look at that!*). For each video, children were asked *Are Sarah and that thing touching? So, did Sarah make that happen or did it just happen on its own?* After seeing both versions, children made two forced-choice decisions, identifying: 1) where Sarah and the object were touching, and 2) where Sarah made the event happen. Positive and negative versions (i.e. where Sarah didn't touch the toy) of these questions were counterbalanced. The pretest procedure was

Event	Agent's action	Outcome effect
(1)	Puppet hops over to land on green squeaky toy	Toy squeaks
(2)	Puppet places ball on ramp, which rolls down to plastic donut	Plastic donut "boings"
(3)	Puppet bends over and places her head on box	Wand on top of box lights up and blinks
(4)	Puppet pushes balanced pendulum	Pendulum tips over and swings
(5)	Puppet wiggles down to globe	Globe lights up and spins around
(6)	Puppet slides over to window shade	Window shade pops up

Table 1: Novel events used for training (events 1-2) and verb-learning trials in Experiment 1. Causal and non-causal versions were created by varying the final position of the puppet and the relative timing of the two sub-events.

then repeated with the second novel event. Children who could not provide correct answers to 3 out of the 4 total forced-choice questions in the pretest were not included in the analysis. Three children were replaced for this reason.

For the critical novel verb test, children saw a trial for each of the four remaining novel events. In each trial, children saw the no-gap (causal) version of the event on one side of the screen, and the gap (noncausal) version on the other. The trial order, as well as version and side presentation for each trial, was randomized for each child.

On each trial children saw the two contrasting movies in sequence, with the same voice-over for both. The voiceover used the target novel verb in intransitive sentences:

Look over here! The tall thing is meeking, it's meeking. Whoa! Watch one more time, it's gonna meek...Wow!

Children were then reminded that *In one movie Sarah made it happen, and in one movie she didn't*. They saw each event a final time, and then the final freeze-frames for both movies were presented. Children heard two test prompts (Positive- *Can you find the movie where she meeked the round thing?*; Negative - *Can you find the movie where she didn't meek the round thing?*) with order randomized across trials. The experimenter waited until the child pointed to a movie; no additional verb prompts were given, but children were invited to 'pick just one movie' if they hesitated to choose. As a manipulation check children were then asked to identify *the movie where they're touching*.

## RESULTS

The dependent measure of interest was how often children chose gap or no-gap versions (i.e. spatiotemporal continuity disrupted or preserved) of each event following different prompts. We predicted that children would choose the no-gap version of events when asked to find the movie where the puppet and toy were touching, and when given positive transitive prompts (e.g. *Where she wugged the round thing*). Note that this sentence suggests a causal referent *only* if children already expect transitive sentences to refer to this kind of scene.

Children's performance was converted to a score between 0 and 4, reflecting the number of trials on which they chose the no-gap (causal) scene. Figure 2 plots children's responses to the three prompt types in Experiment 1. The manipulation check confirmed that children were successfully identifying the scenes

where the puppet and the object were touching; children identified the correct movie at a rate significantly above chance (Wilcoxon signed rank test,  $p < 0.002$ ; 3.04/4 mean correct choices.)

For the positive transitive prompts, the distribution of these scores was also significantly above chance (Wilcoxon signed rank test,  $p < 0.001$ ; 3.08/4 mean causal choices); no children chose fewer than two causal scenes in response to a positive prompt. To show that these choices did not result simply from a global preference for the causal movies, children's responses to negative transitive prompts (*Can you find the movie where she didn't wug the round thing?*) were also analyzed. For these prompts, children's scores were significantly below chance (Wilcoxon signed rank test,  $p < 0.001$ , 0.88/4 mean causal choices); no children chose more than two causal scenes in response to a negative prompt. Patterns were qualitatively

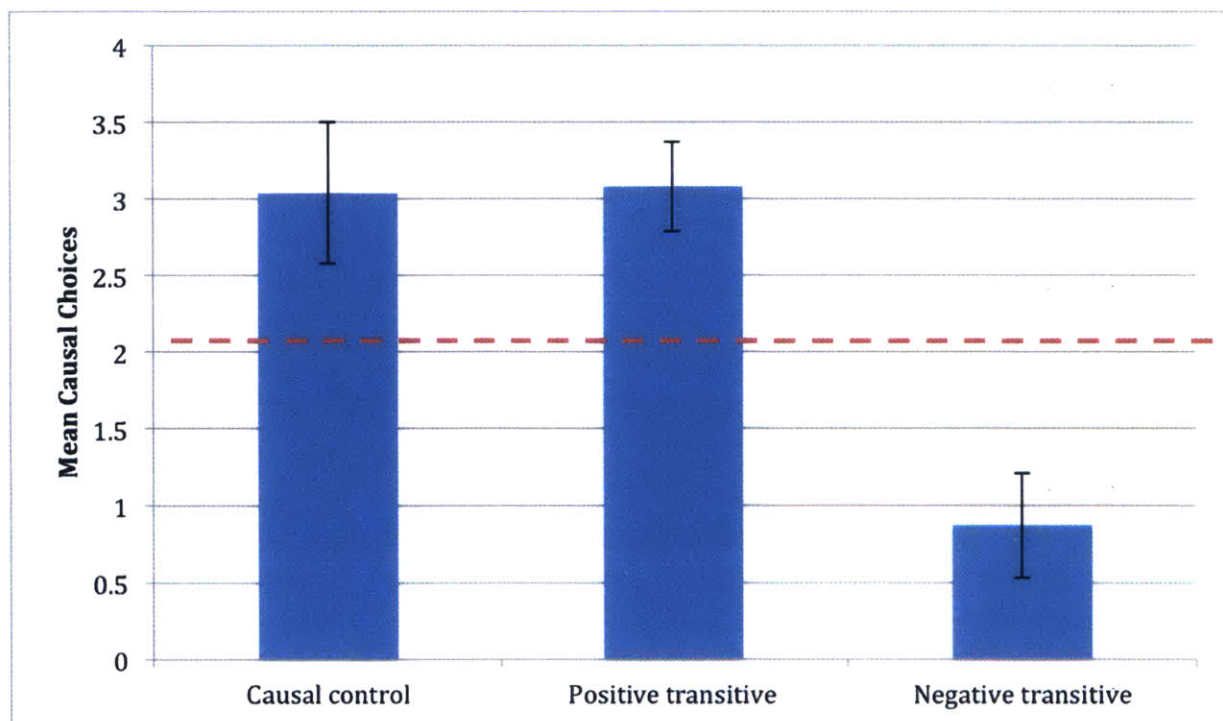


Figure 2: Children's choices of Causal ("No-gap") vs. Non-causal ("Gap") scene variants following the three prompts in Experiment 1. Children heard either a causal control measuring children's basic understanding of the scene ("Find where they're touching"), a positive transitive novel verb ("...where she VERBED the toy", or a negative transitive novel verb ("...where she didn't VERB the toy"). Error bars represent bootstrapped 95% confidence intervals.



similar and there were no significant differences between the performance of 3 and 4 year olds (Wilcoxon signed rank tests, Control prompt:  $p=0.31$ , Positive prompt:  $p = 0.11$ ; Negative prompt:  $p = 0.21$ ).

## DISCUSSION

Experiment 1 indicates that children's interpretation of transitive verbs draws on the cognitive capacities the use for detecting causation in nonlinguistic contexts. Specifically, children's verb learning is sensitive to the causal structure of events, and not only to coarser contrasts between event types such as the number of active participants. 3- and 4-year-olds used spatiotemporal cues to causation to determine the meaning of a novel transitive verb: *Sarah wugged the round thing* led to choices of 'no-gap' scenes with the intact spatiotemporal continuity characteristic of causal scenes, while *Sarah didn't wug the round thing* led to choices of 'gap' variants with timing and contact cues disrupted. Unlike previous novel verb studies, all other properties of the causal and noncausal videos were matched: the participants, the actions performed by the agent, and the physical outcomes were identical in both versions. Thus children must have been relying on the spatiotemporal relationship between sub-events. . Rather than expecting the mere presence of particular kinds of sub-events, transitive verbs led children to expect those sub-events to stand in particular spatial and temporal relation to each other.

The results of Experiment 1 are compatible with the proposal that children have a specific mapping between transitivity and causation. However, there are other interpretations of this finding. First, the presence of the spatiotemporal gap could have led children to interpret the 'gap' movies as involving two separate, sequential events (Puppet's action, object's action) rather than a single event involving both entities. Second, physical contact was used as an index of causation in this study. In English, many verbs of physical contact appear in transitive sentences (e.g. touch, pat, rub). Thus it is possible that children succeeded in Experiment 1 by mapping transitivity to contact rather than cause. As we noted earlier, prior studies have found that children can map transitive verbs to canonical (noncausal) contact scenes when they are pitted against parallel actions. In fact, under some conditions children show no looking-time preferences between the canonical causal and contact scenes given a transitive verb (Naigles & Kako, 1993). Thus, while children's initial nonlinguistic understanding of causation may be closely linked to physical contact (see also Muentener & Carey, 2010), it is important to understand whether children's novel-verb preferences track with causal relations *independent* of physical contact.



Figure 3: Schematic of an event contrast used in Experiment 2. Each novel event was filmed in a causal version and an “agent-last” control which varied the relative timing of the agent’s action and the beginning of the outcome effect.

To address these questions, we return to the literature on causal perception for an additional spatiotemporal index of causation. In addition to spatiotemporal *continuity*, young infants are sensitive to spatiotemporal *order* – if the second billiard ball’s motion starts before the first one arrives to hit it, the causal illusion is broken (Leslie & Keeble, 1987; Michotte, 1963). In Experiment 2, we test children’s preferences for novel transitive verbs with stimuli that hold contact and the closeness of timing constant, but manipulate order.

## EXPERIMENT 2 – TEMPORAL ORDER

Experiment 2 tests a different event contrast: spatiotemporal *order* instead of spatiotemporal *continuity*. As in Experiment 1, each event has a causal version and noncausal version which have the same participants and the same sub-events. In Experiment 2, the spatial relationships in the two versions are also matched, and only the order of the two sub-events differs. In the *agent-first* versions, the agent makes a gesture, and 1 second later (at the endpoint of the actor’s gesture), an outcome effect (e.g. lighting up, spinning around) begins in a novel toy. In the *agent-last* variants, the outcome effect begins (apparently

spontaneously), and 1 second later the agent makes the same gesture. The timing of the causal and noncausal event variations used in Experiment 2 is illustrated in Figure 3.

In order to dissociate causal interpretations from contact interpretations, all events used in Experiment 2 involve ‘action at a distance.’ These scenes can be understood as instances of a causal illusion, like the one that you might experience if you drop a book on a table at the exact moment that someone else turns the lights off. Despite the fact that the agent never touches the object that they change, adults who saw the agent-first videos interpreted them as causal, presumably because the outcome followed immediately on the agent’s action. To ensure that the agent-last versions appeared sufficiently natural, all of the actor’s gestures in Experiment 2 were plausible responses to an interesting event, such as pointing or clapping. By removing the spatial connect between sub-events, Experiment 2 goes beyond previous novel verb studies and allows us to ask whether children associate transitive syntax with events that involve causation but *not* contact.

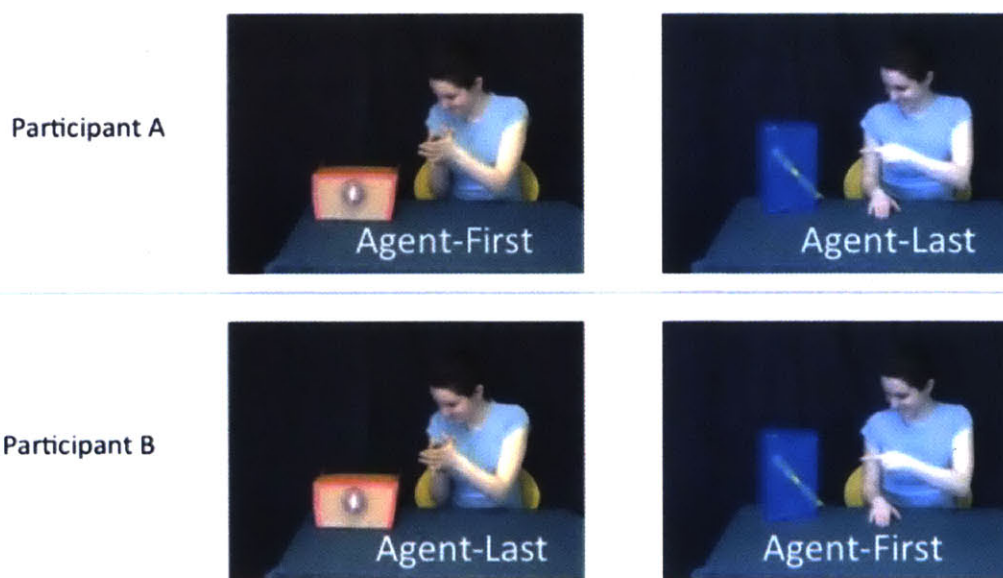


Figure 4: Example of the randomization scheme used in Experiment 2. One participant might see the causal (agent-first) version of the globe event and the non-causal (agent-last) version of the spinner event, while another child would see the reverse. In the Transitive condition, both would then be asked to identify “where she’s meeking something”.

This new manipulation required an additional change in our experimental design. In Experiment 1, there was a clear difference in the two scenes, which was visible at the moment when children were asked to select that correct event (i.e. the presence or absence of the physical gap). This would not have been the case for the timing contrast in Experiment 2 – at the end of both the causal and the noncausal events, the actor has executed her gesture and the novel effect has taken place (see Figure 3). To address this issue, children were given stimuli in contrasting pairs (e.g., *She's meeking it* could refer to either the causal version of Event 7 or the noncausal version of Event 8, with event versions randomized between children.) The event randomization scheme is illustrated in Figure 4. In a between-subjects design, we compared scene choices after transitive sentences (*She meeked the toy*) to two control conditions, intransitive sentences (*The toy meeked*), and sentences with no novel verbs which directly probed causal knowledge (*She made something happen*.)

The dependent measure of interest was how often children chose the agent-first or agent-last scenes depending on the prompt type that they heard. Under the specific-mapping hypothesis, we should expect that children who hear a transitive prompt (*Find where she VERBED something*) will be more likely to choose the causal (agent-first) version of the test events. In contrast, children who hear intransitive prompts (*Find where something VERBED*) should have no such preference. Under the broad-mapping hypothesis, both agent-first and agent-last versions would qualify as two-participant events, and thus children should have no preference in the transitive condition as well. Finally, the performance of children who were given the causal-knowledge prompt (*Find where she made something happen*) will allow us to validate our manipulation (by showing whether children view the agent-first scenes as more causal) and to determine the degree to which performance in the transitive condition matches or departs from children's ability to report the causal relationships in the study paradigm.

## **METHOD**

**Participants:** Preschoolers were recruited from a local children's museum (n=72 mean age: 4;0, range 3;0-4;11, 36 girls), and tested in one of three conditions, Transitive (mean age 4;1, 12 girls), Intransitive (mean age 4;0, 12 girls), and a Causal Knowledge control (mean age 4;0, 12 girls). Participants were replaced if they were unable to reach criteria on the pretest training (n=20.) Nine additional children were replaced due to refusal to participate or parental interference. All children received a sticker and award certificate for their participation at the end of the session.

**Materials:** Agent-first and agent-last versions were created for four new novel events. All events involved a female actor interacting with a novel object. In contrast to the events used in Experiment 1, the agent did not touch the novel objects, but made gestures toward them. The sub-events of each novel event are shown in Table 2. In the agent-first version of each movie, the initiation of the apparatus' effect was closely timed to follow to experimenters' gesture. The agent-last version of each movie was created by filming a version in which only the timing was altered: the actor began her gesture 1 second after the apparatus activated. Because all the actor's gestures were chosen to be plausible 'social responses' to an interesting effect, adults found these movies natural, but did not view the actor as playing a causal role.

In all videos, the event was played through three times, ending on a still shot showing the result and the final position of the experimenter. Videos varied between four and five seconds in total length and the agent-first and agent-last versions of each event were equal in length. In addition to these stimuli, two additional causal movies and two additional 'social response' movies were used during the pretest trials. Videos of the experimental stimuli are available at <http://web.mit.edu/mekline/www/Spatiotemporal/>.

Video stimuli were presented in the same manner as Experiment 1, and the helicopter toy was also used for the warm-up phase of Experiment 2.

**Procedure:** As in Experiment 1, each session consisted of an introduction, a pretest and the main novel verb test. During the introduction, children were introduced to 'my friend Sarah', a puppet who liked to

Event	Agent's action	Outcome effect
(7)	Actor claps while looking at globe box	Globe lights up and spins around
(8)	Actor points at windmill box	Windmill spins
(9)	Actor slaps the table	Balanced pendulum tips over and swings
(10)	Actor raises both hands toward herself	Window shade pops up

Table 2: Novel events used for verb-learning trials in Experiment 2. Causal and non-causal versions were created by varying which of the two sub-events began first. Note that while outcome effects are repeated from the novel events (1)-(6) used in Experiment 1, the spatiotemporal relationships between agent actions and outcomes differ (see Figure 3).

say silly words. After this the experimenter showed them the helicopter apparatus, demonstrating that *Sometimes, Sarah makes it happen, like this...But sometimes, she just watches it happen, it happens on its own because there's a battery inside*. Then the Sarah puppet either approached the toy without touching it, and then the helicopter activated, or she approached it after activation. Note that the causal demonstration for Experiment 2 was different from the one for Experiment 1 in two critical ways: the toy activated *immediately* when Sarah approached and Sarah never touched the toy. Thus the perception of cause was driven by contingency rather than contact. After this demonstration, children were prompted to activate the toy, and shown again that it could activate spontaneously. Then the experimenter prepared children for the rest of the session by explaining that in the movies they would see, *Sometimes, my friend Hannah makes something happen. And sometimes she just watches it happen – it happens on its own because there's a battery inside*. Again, as in Experiment 1, children never heard any events described with transitive sentences during the warm-up.

Children then moved to the video presentation, beginning with two pretest trials which used the training movies described above. Each pretest trial consisted of one agent-first and agent-last movie. The child viewed the two videos, one at a time, and was asked after each video whether the actor *made it happen or watched it happen*. Finally the children were asked to *Find where she made it happen* and *Find where she watched it happen*. Children who could not provide correct answers to 3 out of the 4 total forced-choice questions in the pretest were not included in the analysis. 20 children were replaced for this reason.

During the novel verb test, children saw two trials. The trial order, as well as version and side presentation for each trial, was randomized for each child. In each trial, children saw the agent-first version of an event (e.g. Event 7) on one side of the screen, and the agent-last version of a *different* event (e.g. Event 8) on the other. On each trial, children saw the contrasting movies presented sequentially, accompanied by identical, neutral voiceovers (*Look over here, look at that, wow!*) Children were then reminded that *In one movie Sarah made it happen, and in one movie she didn't*. Finally, children watched both movies playing simultaneously a final time, accompanied by a voiceover appropriate to the between-subjects condition:

Transitive: *She's gonna meek something. She meeked it! Wow, she meeked it!*

Intransitive: *Something's gonna meek. It meeked! Wow, it meeked!*

Causal Knowledge: *She's gonna make something happen. She made it happen! Wow, she made it happen.*

The final freeze-frames for both movies persisted, and children were asked to select the movie *where she fooped it/where it fooped/where she made something happen*. The experimenter waited until the child selected a video before continuing, providing only general prompts (*Go ahead and pick!*) if the child did not immediately point.

## RESULTS

Our analyses focused on how often children chose agent-first or agent-last versions of the events and whether this depended on the type of prompt that they heard. The dependent measure was the number of test trials (out of two) on which the child selected the agent-first scene. The results of Experiment 2 are summarized in Figure 5. Children who heard the causal control questions selected the agent-first scenes at a rate above chance but below ceiling (Wilcoxon signed rank test,  $p = 0.022$ ; 1.38/2 mean causal choices), indicating that the contingency was used to infer a causal link between the two events but that either the inference was difficult for children to make or the task itself was demanding. Children who heard

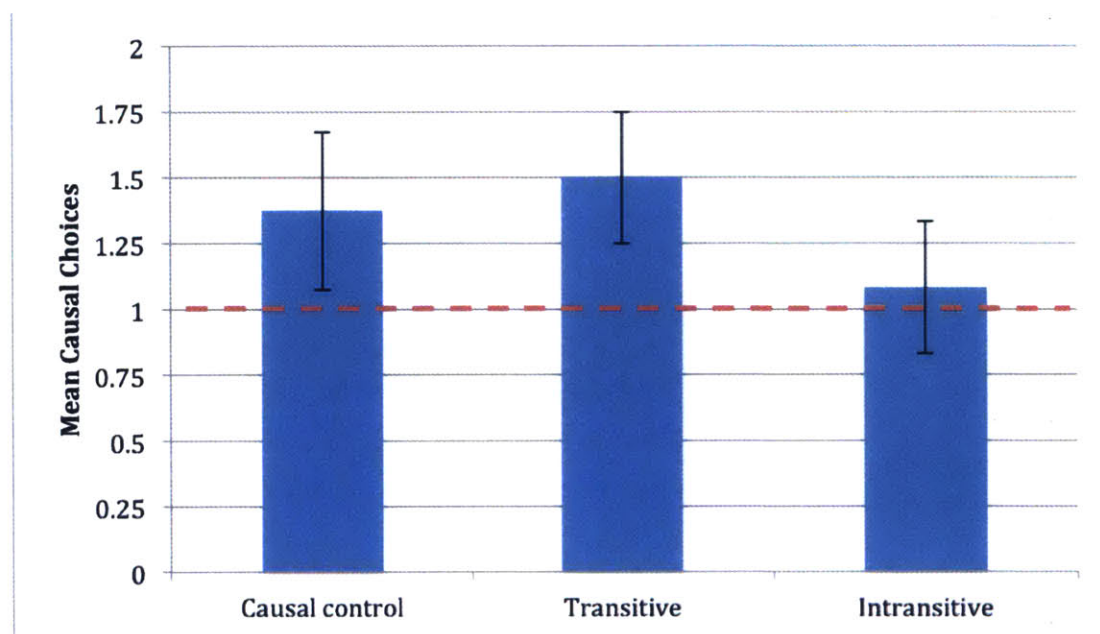


Figure 5: Children’s choices of causal vs. non-causal (agent-first versus agent-last) scene variants following the three prompts in Experiment 2. Children heard either a causal control measuring children’s basic understanding of the scene (“Find where she made something happen”), a positive transitive novel verb (“...where she VERBED something”, or an intransitive novel verb (“...where something VERBED”). Error bars represent bootstrapped 95% confidence intervals.

transitive prompts, also showed a significant preference for the agent-first scene (Wilcoxon signed rank test,  $p < 0.002$ ; 1.5/2 mean causal choices). This was not the case for children who heard intransitive prompts (Wilcoxon signed rank test,  $p = 0.53$ ; 1.08/2 mean causal choices).

In addition to differences from chance, we also performed planned comparisons of the Transitive versus Intransitive conditions, and of the Transitive versus Causal Knowledge conditions. Children in the Transitive condition were more likely to choose agent-first scenes than the children in the Intransitive condition (Wilcoxon signed rank test,  $p = 0.02$ ). Performance in the Transitive and Causal Knowledge conditions were not statistically different from one another (Wilcoxon signed rank test,  $p = 0.61$ ).

There were no significant differences between the performance of 3 and 4 year olds, and patterns were qualitatively similar across the three conditions (Wilcoxon signed rank tests, Transitive prompt:  $p = 0.09$ ; Intransitive prompt:  $p = 1$ , Control prompt:  $p = 0.41$ ). Three-year-olds in the Transitive condition chose causal scenes somewhat less often than four-year-olds (1.31/2 vs. 1.73/2 mean causal choices); however, this was also true for the Causal Knowledge control condition (1.23/2 vs. 1.55/2 mean causal choices), indicating that this difference had to do with relative success at identifying the causal relation or proficiency on the task in general.

## DISCUSSION

In Experiment 2, children showed a clear preference to map new transitive verbs to causal scenes, rather than closely-matched noncausal foils. Specifically, children who were asked to *Find where she VERBED something* consistently chose events where the agent moved before the onset of the outcome action, rather than where she moved directly *after* its onset. This preference was specific to the transitive structure: Children who heard *Find where something VERBED* did not show this pattern. The performance of children learning transitive verbs was also very similar to children who were asked to find *Where she made something happen*, suggesting that their imperfect performance in the transitive condition was primarily due to uncertainty about the causal structure or lapses in attention, rather than uncertainty about the mapping between transitive verbs and causal scenes.

These results indicate that children are able to take fine-grained spatiotemporal information into account when determining the meaning of a novel verb. The causal and noncausal stimuli in this experiment were constructed to hold timing (as well as participant and sub-event information) constant – in both types of movies, one sub-event was initiated one second after the other. Knowing that an intentional action by an agent can have causal power is not sufficient – only when the agent's action preceded the physical



outcome (rather than vice versa), did children interpret the scene as causal and map novel transitive verbs to that scene.

## GENERAL DISCUSSION

Across two experiments, children aged three and four years old had a bias to map transitive syntax to causal scenes. Unlike all previous studies of this type, we manipulated specific spatiotemporal cues to causation, rather than using contrasting prototypes from two different event categories. These studies reveal that children's expectations about transitive verbs map closely to their nonlinguistic understanding of causal events: even when two alternatives have identical sub-events, and differ only in the spatial and temporal relations between those sub-events, children expect transitive syntax to refer to scenes where the agent's sub-event is a plausible cause of the outcome. In Experiment 1, children expected transitive novel verbs to refer to spatiotemporally intact causal interactions rather than to scenes with a spatial gap and temporal pause between the agent's action and the outcome event. In Experiment 2, they preferred scenes where the agent's action closely preceded the activation of the toy over identical scenes in which the timing of these two events was switched. This was true even despite the fact that there was no physical contact between the agents and the toys.

This work represents an important advance in our understanding of how young children map between syntactic structures and semantic event representations. Our results strongly suggest that children have a mapping between a fairly abstract notion of causation and the transitive construction, and rule out a number of alternatives. 3-4-year-olds' preferences cannot rest solely on properties either of the agent's action or particular outcome effects, although these may serve as important cues during naturalistic verb learning. In distinguishing between possible referents of a new transitive verb, children attend to exactly those fine-grained cues that drive their awareness of causal events from infancy. This suggests that, by the preschool years, children possess at least one specific mapping between transitive syntax and causal events, rather than only a broader mapping to all (asymmetric) two-participant events.

This finding makes predictions about a range of other manipulations that should affect transitive verb learning. For example, we should expect that children will be able to systematically link transitive verbs to causal events even when there is no spatial or temporal information in any individual event that reveals its causal structure. Gopnik and colleagues (2004) have demonstrated that children can use patterns of covariation to determine which of several possible causes is responsible for an effect (using the 'blicket detector' paradigm). If children interpret transitive verbs as encoding cause, then they should assume that

a novel transitive verb picks out the event with the statistically probable cause even if no spatiotemporal cues to causation are present in the events that are currently being labeled.

In the remainder of the discussion, we address possible alternative interpretations of these studies, implications of these findings for preschoolers' knowledge of other transitive verbs, and the possible developmental origins of these mappings.

#### *ALTERNATE EXPLANATIONS*

One alternate interpretation of Experiment 1 is that children viewed the 'gap' movies as two separate, sequential events (puppet's action, object's action) each with a single participant, rather than as a single event (with two subparts) that involved both entities. Event segmentation is an important and difficult problem (Zacks, 2010). Some degree of event segmentation (e.g., picking out a spatiotemporal event chunk) would seem to be a pre-requisite for evaluating possible conceptualizations of that event. Furthermore, we know that spatiotemporal contiguity is likely to play a role in children's event segmentation: young children look longer to videos where pauses are inserted in the middle of actions, than to videos where the pauses coincide with event boundaries (Baldwin, Baird, Saylor, & Clark, 2001). In fact, it is possible that the process of segmenting events plays a role in the discovering their causal structure: sub-events which are close in time and space, and thus which are segmented as a single event, are presumably more likely to be causally linked as well.

#### *LEARNING NONCAUSAL VERBS*

While a specific mapping between causation and transitive syntax gives children an advantage for learning some verbs, it should be a disadvantage for learning others. In particular, our findings raise questions about how children learning languages like English acquire noncausal transitive verbs like *touch*, *love*, or *see*. The agent-last versions of the events in Experiment 2 can be naturally analyzed as the agent's response to the interesting toy. Children might thus have understood the novel transitive verbs as social or emotional response verbs (e.g. *She applauds the performance.*) Instead, they mapped the verb to events with the spatiotemporal structure of physical causal relationships. This suggests that having a strong bias to map transitive syntax to causal meanings might help children in some cases, but hurt them in others. Just as a bias to interpret novel nouns as labels for whole objects could make learning other kinds of nouns more challenging (Markman, 1992), a strong causal-transitive bias might make interpreting other kinds of transitive verbs difficult. In the case of transitive verbs, there is evidence that English-learning preschoolers struggle with some classes of noncausal transitives. In particular, English

has two classes of emotion verbs: verbs where the subject is experiencer of the emotion (e.g. the lion **fears** the tiger), and verbs where the subject is the causer of the emotion (e.g. the tiger **frightens** the lion)(Pesetsky, 1995). Despite the fact that (noncausal) *fear*-type verbs are more common by token frequency, children appear to have more difficulty interpreting them correctly (Hartshorne, Pogue, & Snedeker, in press). Nevertheless, children know and use many noncausal transitives by the time they are three years old (Fenson et al., 2000). These might be mapped to event representations in several ways. First, children might have a broad two-participant schema available alongside the more specific and salient causal mapping. Second, they might have additional narrower-range event schema mappings for other classes such as contact or emotion verbs. In either case, the results of the present studies suggest that children might need additional evidence to rule out a privileged causal-transitive mapping.

While some of this evidence could come from observing a range of scenes in which the verb is used, children might be able to use the distribution of syntactic frames to narrow in on the meanings of transitive verbs. Most verbs appear in more than one syntactic frame, and the combination of frames is likely to be more informative about its meaning than any single frame (see e.g., Fisher, Gleitman, & Gleitman, 1991; Levin, 1993). For example, the causative/inchoative alternation (e.g. She blicks the toy/the toy blicks.) is typically used for events that involve a change of state or location and that can be caused by an external agent. Causal transitive verbs in English generally participate in this linguistic alternation, while other transitives do not. For instance, while "She touches the lamp" is acceptable in English, "The lamp touches" is not an acceptable under the reading where the lamp is the affected entity.

If children understand the additional information carried by this alternation, then the failure of a particular verb to participate in it might serve as a (probabilistic) source of evidence for a noncausal meaning. Indeed, several novel verb studies have shown that children prefer different kinds of event prototypes following different syntactic alternations (see e.g. Fernandes, Marcus, Di Nubila, & Vouloumanos, 2006; Naigles, 1996, 1998; Scott & Fisher, 2009). Nevertheless, the present results suggest that children's causal biases go beyond a specific connection to just those transitive verbs that participate in the inchoative alternation. In Experiment 1, but not in Experiment 2, children were given evidence that the novel verb participated in this alternation. Thus their tendency to interpret the transitive verb as causal could reflect knowledge of the alternation, in addition to a mapping for the frame itself. Critically, however, the results of Experiment 2 demonstrate that 3-4 year olds associate transitive syntax with causation, even *in the absence of* evidence for the inchoative alternation. This is consistent with the cross-linguistic connection between transitivity and causation and the tendency for adult (English speakers) to

interpret novel transitive verbs as causal (Kako & Wagner, 2001; Kako, 2006; Levin & Rappaport Hovav, 2005).

### *DEVELOPMENTAL TRAJECTORY*

Our findings also raise the question of how a link between cause and transitivity emerges during development. Complicating this question is the fact that children are not born with fully adult-like causal knowledge. While spatiotemporal cues to causation affect infants' attention from the first year of life, children initially have very different expectations than adults about the causal powers of animate and inanimate entities (Leslie & Keeble, 1987; Muentener & Carey, 2010). Any theory of early connections between argument structure and causation will have to account for how children come to associate linguistic representations with cognitive representations that are themselves still developing.

We see three possible accounts of how the observed mapping between cause and transitivity could develop. First, as proposed by Fisher and colleagues, younger toddlers may begin with a global bias to match the number of linguistic arguments to the number of event participants (leading to a broad 'two participant' preference for transitive verbs.) As they learn their native language and as their nonlinguistic cognition continues to develop, additional, more specific biases (like those observed in the present studies) might arise, to the degree that they are supported by evidence from the particular language the child is learning.

Second, like preschoolers, infants may have a specific bias to map transitive sentences to causal scenes. To the best of our knowledge, there is no existing work that clearly demonstrates that this bias exists in younger children. While prior work suggests that children as young as 19-21 months prefer to map transitive verbs to prototypical causative scenes rather than other scene categories (Arunachalam et al., 2013; Yuan et al., 2012), the scene contrasts used in these studies vary along multiple dimensions leaving the conceptual basis of the preference unclear. There are, however, good reasons for supposing that a causal-transitive bias might be present early in life. As we noted earlier, while causal knowledge develops throughout life, some of the guiding principles of causal reasoning are in place by 6 months of age (Leslie & Keeble, 1987). Furthermore, the causal-transitive mapping is cross-linguistically robust, raising the possibility that it has origins in conceptual and learning biases that young children bring to the problem of language acquisition (though see Christiansen & Chater, (2009) for other explanations of cross-linguistic patterns). This hypothesis would be particularly compatible with theories where causation plays a central role in argument structure (Croft, 2012).

A third possibility is that toddlers begin with an expectation that the arguments of a transitive sentence will have *asymmetric* roles, and that the roles of Causer and Cause-ee are particularly good exemplars of this asymmetry. For example, Dowty (1991) proposes that verb argument structure is guided by a set of criteria that define prototypical agents and prototypical patients. The argument with the most proto-agent properties becomes the subject of an active sentence, while the argument with the most proto-patient properties becomes the direct object. By Dowty's criteria, the two arguments of a causal verb are maximally suitable for a transitive construction; the Causer has most of the prototypical agent properties and the Cause-ee has most of the prototypical patient properties, resulting in maximally asymmetric roles. In contrast, the asymmetries in events involving contact or emotional state are less pronounced. This might explain the distribution of transitive verbs across the world's languages, with verbs for causal events, but not other types, appearing reliably with transitive syntax (Levin & Rappaport Hovav, 2005). It is also congruent with the patterns of attention that young infants display: 6-month-olds are more surprised when the (billiard ball) participants of a causal interaction change than when the participants of spatiotemporally altered sequences change (Leslie & Keeble, 1987).

Whatever the structure of children's initial expectations about how events map to language, the fact that individual languages differ in the range of events that get described by transitive syntax indicates that language learning involves changes in these mappings over time. Understanding how children represent events (i.e. how they define the concepts of 'agents', 'roles' and 'asymmetry') during language acquisition will be critical to understanding how their theories about verb meaning and argument structure develop. The methods presented in this study provide an important avenue for testing these questions.

## **CONCLUSIONS**

These experiments are the first novel-verb comprehension studies to bring children's semantic knowledge into direct contact with their nonlinguistic causal knowledge. By the preschool years, rich and structured representations of causal events are recruited for verb learning. Specifically, we show that children use syntactic information to guide inferences about transitive verb meaning that are closely related to their nonlinguistic concepts of causation. This finding opens a wide avenue for exploration into the relationship between children's linguistic and non-linguistic knowledge. If we can clarify how children's inferences about word meaning make contact with their nonlinguistic representations, then we will be better equipped to understand how children learn about the world from second-hand testimony, and update their beliefs about world from the sentences they hear (Bonawitz et al., 2010; Harris & Koenig, 2006; Harris, 2002; Nazzi & Gopnik, 2001; Schulz et al., 2008). Critically, this study shows that any examination of the

semantics of early language must be considered a question of cognitive development as well as linguistic development. Understanding early representations of verb argument structure will require understanding how children in the first few years of life are representing the scenes they see. Even as infants, children have rich, but not necessarily adult-like, representations of what constitutes an agent, a cause, or an event (C. Baker et al., 2011; Luo et al., 2009; Sobel et al., 2004; Spelke, 1990; Woodward, 1998). Any detailed understanding of what children encode in their early verb meanings must reckon with the *kinds of meaning* that a toddler might have available to encode in language. Exploration into children's early syntactic representations can therefore become a critical part of the effort to understand how humans across the lifespan represent events (Baldwin, Andersson, Saffran, & Meyer, 2008; Baldwin et al., 2001; Wolff, 2008; Zacks, 2010). By bringing together the linguistic tests for novel verb comprehension with stimulus manipulations from research on prelinguistic cognition, we can make detailed, testable predictions about how children make inferences about language from the events they see, and how language in turn reflects the structure of event representations.

## CHAPTER 3 – DO WORD ORDER PATTERNS IN GESTURE REFLECT GENERAL COMMUNICATIVE PATTERNS OR MODALITY-SPECIFIC PRODUCTION CONSTRAINTS?

### ABSTRACT

A fundamental typological variation in the world's languages is their basic word order; most spoken languages are either SOV or SVO. Previous work has related this typological pattern to a striking finding in ad-hoc gesture production: adults across a wide range of languages tend to gesture events with inanimate patients in the order SOV and those with animate patients in SVO order (Gibson, Piantadosi, et al., 2013). This has been interpreted as support for a noisy channel model of communication under which producers attempt to reduce ambiguity for comprehenders. However, it has been pointed out that this pattern might also result from the particular kinds of gestures that people tend to use for different kinds of event participants (Hall, Mayberry, & Ferreira, 2013). In two gesture production studies and an accompanying comprehension study with English speaking adults, we tested whether different kinds of modifications would affect the word orders that participants used in the ways predicted by these two hypotheses. The first kind of modification was to the information carried by the gestures (an ad-hoc case-marking system) and was intended to alter only gesture *ambiguity*. The second modification was designed to alter only gesture *form*, and not the ambiguity of the gestures. Two main findings emerged from these

tests: first, simple modifications to the gesture paradigm can have a profound effect on the orders used: both attempts significantly decreased the use of SVO orders, even though participants were gesturing about the same events under the same communicative contexts. Second, the two modifications had more complex effects than anticipated, affecting *both* ambiguity and form such that a unique cause for the dramatic order changes could not be determined. We conclude that gesturing paradigms, while a striking and naturalistic example of ad-hoc communication dynamics, are also affected by the particular modality and as such do not provide straightforward evidence for noisy channel theories of word order typology.

## INTRODUCTION

A fundamental typological variation in the world's languages is their basic word order: in English, most two-participant sentences express, in order, the agent, action and patient (or subject, verb, and object, SVO)<sup>1</sup> of an event i.e. *The boy kicks the ball*. In contrast to English, the subject and object are typically expressed before the verb in Korean (SOV):

(1) 소년이 공을 찬다

sonyen -i kong -ul cha -nta

boy -NOM ball -ACC kick -DEF

*The boy kicks the ball*

Strikingly, although there are six possible basic word orders (SVO, SOV, OSV, OVS, VSO, and VOS), the large majority of languages are either SVO like English (41%) or SOV like Korean (47%) order (Dryer, 2005).

Evidence from signed languages, patterns of language change, and gesture-communication experiments suggests that one word order, SOV, might be somehow primary or default in cognition and language (Givón, 1979; Goldin-Meadow & Feldman, 1977; Goldin-Meadow, So,

---

<sup>1</sup> Throughout this paper, we use the S,O,V notation to refer to agents, patients, and actions of



Özyürek, & Mylander, 2008; Newmeyer, 2000). In many emerging sign language communities (formed by deaf individuals living in the same community but without access to an existing signed language), SOV word orders (and related SV and OV orders) are common even when surrounding spoken or signed languages use other orders (Meir, Sandler, Padden, & Aronoff, 2010). Similarly, a survey of 42 existing sign languages found that SOV order is always permissible; the same is not true for SVO order (Napoli & Sutton-Spence, 2014). In the history of spoken languages, SVO languages have sometimes evolved out of SOV order languages, but the reverse does not seem to be the case (Gell-Mann & Ruhlen, 2011). This suggests that SOV may have been the initial or preferred form for spoken human languages.

In a controlled gesture-production experiment, Goldin-Meadow et al. (2008) reported a cross-linguistic SOV word-order bias in a paradigm where participants are asked to convey event meanings using gestures. Critically, these participants spoke a variety of native languages with different basic word orders, but did not know any signed languages. In this experiment and in several replication studies, participants tend to gesture the agent, the patient, and then the action for canonical transitive scenes with an animate agent and an inanimate patient. For example, when asked to gesture the meaning of an animation depicting a boy kicking a ball, participants typically gesture *BOY BALL KICK*, regardless of whether they speak a verb-final language like Korean, Turkish or Japanese, a verb-medial language like English, Spanish, Italian, Chinese, or Russian, or a verb-initial language like Tagalog or Modern Irish (Futrell et al., 2015; Gibson, Piantadosi, et al., 2013; Goldin-Meadow et al., 2008; Hall, Ferreira, & Mayberry, 2014; Hall et al., 2013; Langus & Nespors, 2010).

Although this basic SOV gesture/sign order is robust across language groups, it can be significantly modulated by the nature of the event to be communicated. In particular, participants appear to be sensitive to the potential *reversibility* of the events, i.e. whether either character could conceivably act as the agent of the action. For instance, the sentence *The boy kicks the girl* can be plausibly reversed to *The girl kicks the boy*; the same is not the case for a sentence with an inanimate object like *The boy kicks the ball*. For speakers of SVO and VSO languages, the default verb-final gesture order (SOV) tends to change to verb-medial (SVO) when participants gesture the meanings of reversible (animate-patient) events (English: Gibson, Piantadosi, et al.,

2013; Hall et al., 2013; Meir, Lifshitz, Ilkbasaran, & Padden, 2010; Russian, Tagalog, Modern Irish: Futrell et al., 2015). In predominantly SOV languages like Japanese, the same animacy-dependent shift from SOV to SVO is observed for more complex events, where the critical transitive event is embedded inside another event, such as *The grandmother said that the boy kicked the girl* (Gibson, Piantadosi, et al., 2013)<sup>2</sup>.

#### *NOISY-CHANNEL MODELS OF WORD ORDER*

Two basic classes of hypotheses have been proposed to explain this animacy effect. The first class of explanation for the SOV/SVO flip focuses on the suitability of different gesture orders for successful communication or recovery of information. Under a *noisy-channel* framework (Aylett & Turk, 2004; GP2013; Jaeger, 2010; Levy, Bicknell, Slattery, & Rayner, 2009; Shannon, 1949), this effect is hypothesized to result from a general pressure to communicate efficiently and robustly. Both producers and comprehenders are sensitive to ambiguity in language, and make decisions to account for it in a rational way (Piantadosi et al., 2011; Piantadosi, Tily, & Gibson, 2012).

When the agent is animate and the patient is inanimate, semantic likelihood alone can guide reconstruction of the intended meaning: even a very rare order like KICKING BOY BALL has only one likely corresponding message. When both characters in an event are animate, the potential for miscommunication is higher (because either could have been the agent). Several research groups have hypothesized that an SVO order may be less ambiguous than an SOV order, since the two characters are placed farther apart (GP2013; Fischer, 1975; Meir, Lifshitz, et al., 2010; Napoli & Sutton-Spence, 2014). Specifically, GP2013 argue that the positions of the noun phrases with respect to the verb (before and after) provide a stronger cue about their roles in the event, because SVO order is more robust to noise. If speakers and gesture producers are

---

<sup>2</sup> Of these studies, this paper will focus specifically on interpretations offered by Gibson, Piantadosi, et al. (2013) and Hall, Mayberry, & Ferreira (2013). These studies will be referred to as GP2013 and H2013 for simplicity.

sensitive to these kinds of information concerns, they may be implicitly designing their gesture sequences to be robust to noise.

Of course, the robustness of this SVO order for communication depends on what kinds of ‘noise’ pressures might affect it. In the noisy-channel framework, noise can include any factor (modality-specific or modality-independent) that affects either how accurately the producer says what they want to say, how the message might be degraded during transmission, or how accurately the comprehender understands or remembers it (cf. GP2013). What kinds of noise might affect the transmission of the simple 3-symbol sequences in these gesture paradigms? We consider three possible noise models for this task: two Deletion channels and two Swap channels. Two of these models result in more robust transmission of SVO order than SOV order, while two are agnostic.

In a Deletion channel, one of the three symbols may fail to reach the comprehender (e.g. because the producer forgets to produce it or chooses a confusing gesture, or because the comprehender forgets or fails to understand). Depending on how this deletion occurs, the symbol might be lost entirely, or the comprehender might know that some symbol occurred but be unsure of its identity. Examples (2a) and (2b) show possible transformations under these two models.

(2) a. Full deletion

SOV-> SV or OV

SVO-> SV or VO

b. Partial deletion

SOV -> SxxxV or xxxOV

SVO-> SVxxx or xxxVO

In model (2a), SOV order leads to a problem for reversible sentences: if the comprehender receives a message like *BOY KICK*, is the boy the subject or object (agent or patient) of the sentence? For non-reversible scenes, partial recovery of the meaning is still possible – if one of the noun phrases is lost, semantic constraints on possible agents will lead to the conclusion that *BALL KICK* refers to a meaning of a ball being kicked rather than a ball kicking someone. On

the other hand, SVO order is more robust to this kind of noise: the two possible messages are BOY KICK or KICK GIRL, and the position of the noun relative to the verb indicates whether it is the subject or object.

In model (2b), this kind of partial meaning recovery will be equally possible for both SOV and SVO orders: it is possible in either case to determine the relative order of the two nouns, and comprehenders regularly use a subject-first heuristic to assign roles to event participants (cf. Hall, Ahn, Mayberry, & Ferreira, 2015). Which of these models is more plausible? In typical speech, comprehenders are more likely to be unsure of a word than to miss its presence entirely, which would require entirely failing to process a segment of a sentence. However, Gibson, Bergen, et al. (2013) have shown that when English speakers evaluate possible speech errors, they are very willing to assume that a producer has failed to include a word.

In a Swap channel, any two symbols may be transposed with one another (again, this might be due to an error by either the producer or the comprehender). In the first version of the noise channel (3a), we assume that transposition is more likely for adjacent elements than for separated elements, which accords with some types of errors in memory and natural language such as within-word Spoonerisms (Estes, 1972; Henson, 1998; MacKay, 1984). In the second (3b), we assume only that nouns are more likely to transpose with one another than a noun and a verb.

(3) a. Adjacent swap

SOV-> OSV or SVO

SVO-> VSO or SOV

b. Noun swap

SOV -> OSV

SVO-> OVS

In the first version of the Swap noise model, SOV is again less robust to communication errors: the linear order of S and O can be reversed, such that participants will misunderstand the roles of the characters in a reversible sentence. In contrast, the relative order of S and O is always

preserved when the base order is SVO. In model (3b), both SOV and SVO are equally vulnerable to the kind transpositions that will cause comprehender error.

Under a noisy-channel framework, we hypothesize that gesture production patterns, specifically the animacy-contingent flip from SOV to SVO orders, result from producers who assume a noise model like (2a) or (3a). Note that this does not necessarily require that comprehenders actually make these kinds of errors (i.e. are subject to that kind of noise), merely that producers *assume* that these errors are likely. GP2013 also point out that similar patterns may emerge when gesture producers design utterances that are robust for their own memory or internal representations.

#### *PRODUCTION-CONSTRAINT MODELS OF WORD ORDER*

The second kind of explanation for these word order effects involves constraints on the producer that result specifically from the modality of gesture. Hall and colleagues (H2013) conducted a series of experiments very similar to those conducted by Goldin Meadow et al. (2008) and GP2013, focusing primarily on the distribution of non-SOV orders produced for reversible events. In addition to SVO orders, they also found an increased prevalence of two additional orders: OSV(O) and SOSV(O) (a similar distribution was found by Meir et al. 2010, though not by GP2013). These orders accounted for 30% of reversible trials, and are not predicted by noisy-channel or other communicative explanations, because both subject and object appear on the same side of the verb. Instead, H2013 argue that these orders, along with SVO, satisfy a constraint against *role conflict* in reversible sentences, a mismatch that arises between the gestures that a participant produces using themselves (i.e. their own bodies) as part of the sign.

For animate entities in both subject and object roles, participants commonly produce **body-based** gestures (e.g., signing PONYTAIL on their own heads to refer to a girl, or JACKET on their own torso to refer to a person in a jacket.) This is in contrast to gestures made in the signing space in front of the gesturer (e.g. sketching the outline of a jacket.) Importantly, body-based gestures are also used for communicating verbs/actions; when this is done, gesturers almost always take on the role of the agent of the action. For instance, THROW is gestured by raising one's arms in a throwing motion, not by shifting one's body as though thrown by someone else. When gesturing events with inanimate patients, this is relatively straightforward because there is only one human

involved in the event. However, when gesturing events with animate patients (reversible events), role conflict occurs because of shifts in perspective of who the ‘body’ in the gesture belongs to. Thus, participants experience conflict when gesturing an animate *patient* in a body-based way and then moving immediately to gesture a verb from the perspective of the *agent*. According to H2013, producers prefer that the animate entity most immediately preceding the verb should be the agent, so that the body-based gestures match perspective. Thus, participants avoid a local “OV” sequence for reversible events, producing SVO and the other observed orders instead. Critically, this is described as a constraint on production, not an attempt to facilitate communication: the gesture sequence feels unnatural or awkward to the producer, and so they take steps to avoid it.

Hall and colleagues do acknowledge the likely role of communicative pressure on gesture ordering, but argue that the actual challenges that comprehenders face in understanding gesture sequences in a naturalistic task do not follow the predictions of a noisy-channel model (Hall et al., 2015, 2014). In particular, in a comprehension task where participants guessed the meanings of three-gesture sequences, both Korean- and English-speaking participants were very successful at interpreting SOV and SVO orders, even for reversible sentences. For English-speaking participants there was a small performance difference in favor of SVO; in the Korean test population there was no difference, but a small trend in favor of SVO as well. This trend moderately supports a noisy-channel interpretation (e.g. the advantage predicted by noise models 2a and 3a), but Hall and colleagues point out that performance in both SOV and SVO conditions was well above chance. Participants relied on a consistent subject-first heuristic, accurately assigning the agent and patient roles to both gesture orders. In addition, Hall et al. found (by modeling communicative success from the distribution of gesture orders in production and the corresponding success rates in comprehension) that the orders used for non-reversible would have produced *higher* comprehension rates on reversible trials than those actually produced for reversible events (this is due to the occurrence of orders like OSV in the production data for reversible sequences.)

These comprehension results are limited, however, because they present full gesture ‘sentences’ to comprehenders in a paradigm without significant memory or task demands. In particular,

participants were allowed to watch the gesture sequences as many times as they wished. There was thus little opportunity for the kinds of communication breakdown that noisy-channel approaches are intended to model. The existing comprehension results suggest that the mere presence of multiple word orders is not enough to create a significant barrier to comprehension. The noisy-channel framework proposed by Gibson et al. and others has proposed more significant types of noise that probabilistically delete or swap words, representing a relatively large loss of information. Modeling of noise channels like this has been successful in explaining other aspects of grammatical structure and error detection (cf. Gibson, Bergen, & Piantadosi, 2013). However, it remains an empirical question what kinds of noise (either literal or figurative, in the form of memory errors on the part of producer or comprehender) gesturers account for when designing their utterances.

#### *TESTING EXPLANATIONS FOR WORD ORDER IN GESTURE*

Findings in gesture experiments have been taken to be informative about the underlying structure of human language and cognition (cf. Langus & Nespors, 2010). Indeed, the task is relatively natural for participants, and has the significant advantage that participants often do not realize that order is the main aspect of their gestures being studied – anecdotally, in our own studies we have observed participants who believed they had used English word order throughout the task when in fact they had produced significant numbers of SOV sequences. The noisy-channel framework takes gesturing as an analog of how communicative pressures operate in human language; if this is the case, new findings in the gesture paradigm can motivate broader hypotheses about the nature of language and language evolution. On the other hand, if the patterns that have been observed in gesture paradigm are due to modality-specific production constraints (i.e. role conflict), this generalization is unwarranted.

Critically, these two positions make different predictions about the kinds of changes to the gesture production task that should affect performance. If gesture sequences, (particular the animacy-contingent shift from SOV to SVO), are primarily shaped by communicative pressures (i.e. avoiding *ambiguity*), then changing those pressures might change the gesture orders observed. Task instructions which merely modify the particular gestures used should *not* have a significant impact on gesture order. On the other hand, if the gesture orders (or at least the shift

to SVO) are due to the kind of production constraints that H2013 describe (avoiding *role-conflict*), then the reverse is true. Instructions which modify only the communicative pressures are unlikely to have an effect on gesture order, while task changes which require participants to use non body-based gestures may have a significant impact on gesture order because they reduce role conflict.

We present three kinds of evidence to determine whether gesturing patterns are the result of producers using noise models like those described in (2-3). First, we conduct an additional comprehension experiment to determine whether comprehenders under memory demand fail or succeed under the conditions that are predicted by GP2013. In particular, we ask whether SOV sequences are mis-remembered more often than SVO. While producer models do not necessarily match actual comprehender error patterns, this may give insight into whether these particular noise models can potentially make communication more robust.

Second, to test whether shifts in word order result from avoiding *ambiguity* (GP2013), we provided participants with a case-marking system. This allowed them to indicate in their gestures the difference between a subject and an object. If producers are using a noise model as described above, this should eliminate the need for SVO orders, leading to SOV orders for both reversible and nonreversible events. Under a Deletion noise channel, a sequence like BOY<sub>subject</sub> KICK is still partially recoverable to the same degree in SOV order as in SVO order. Under a Swap noise channel, a GIRL<sub>object</sub> BOY<sub>subject</sub> sequence can still be fully interpreted (with the verb in either position). If producers' orders are primarily a result of production constraint (avoiding *role conflict*), the manipulation is not predicted to affect order – so long as participants' gestures for animate characters are qualitatively similar. In order to explain any shift in gesture pattern, it will be necessary to see whether role conflict exists to the same extent before and after the case marking manipulation. We therefore code the particular gestures that participants produce to see whether gesture orders can be accounted for by differences in role conflict.

Third, we conduct another manipulation in the gesture task designed to directly test whether avoiding *role conflict* (H2013) will alter gesture orders. By preventing participants from using the body-based gestures that are believed to create role-conflict, we will be able to examine the orders that participants use when not subject to this production constraint. Under a noisy-channel



framework, the particular form of the gestures used is not predicted to affect word order. However, we must again assure that the manipulation is valid by coding for case marking strategies, which participants sometimes spontaneously generate (GP2013). Together these manipulations may allow us a clearer window on the origin of word order biases and on the relationship between the gesture task and typologies in spoken language.

## EXPERIMENT 1: COMPREHENSION

As discussed above, producers operating under a noisy-channel model of communication might not be successful – they might attempt to account for noise in a way that is in fact not effective for the listener. Nevertheless, understanding how listeners actually comprehend sentences proposed to be better or worse under a specific noise model is a critical test for understanding which if any of these theories is actually operating during communication. Hall et al. (2015) found no significant problems in comprehending reversible SOV sentences: speakers of both English and Korean understood these sentences well above chance. Relatedly, they also found that people did not have accurate models of how another person might perform in the gesture paradigm: when shown an event paired with a confederate gesturing in either SOV, SVO, or OSV order, both English and Turkish speakers gave the highest rating to their own native languages, and their ratings for reversible and non-reversible sentences did not follow the patterns seen in production experiments.

Both of these experiments provide possible evidence against some versions of a noisy-channel model, but they also fail to capture some of its critical assumptions. In the comprehension task, participants could see both possible alternative events (e.g. *Boy kicking girl* and *Girl kicking boy*), and were allowed to replay the gesture sequence as often as they liked. The memory demands on the participants were therefore slight, if participants make particular patterns of errors on some sentences over others, they may fail to emerge if those errors are rare. In the rating experiment, it is difficult to know how the gesture-comprehenders expected orders to shift with reversibility (the critical question for the noisy-channel model), because the ratings they gave for the *non-reversible* gestures did not match the orders that producers typically use (e.g.

high ratings for SOV, lower ratings for other orders). For this reason, it is difficult to interpret their shifts in judgments between non-reversible and reversible scenes.

We can raise the likelihood of memory and comprehension errors in general by requiring participants to complete a concurrent task. This interference paradigm may not match the demands on comprehenders during language comprehension, but can at least establish whether the expected patterns emerge when participants must simply remember and recall gesture sequences.

In this experiment, participants learned gesture sequences and then later verified whether a test sentence matched the gesture sequence they had just seen. We asked participants to remember these gesture sequences while performing a number-sequence memory task. In pilot testing of a gesture-to-gesture matching paradigm, some participants reported that they had tried **not** to think about the meaning of the gestures in order to better remember the specific motions they were seeing. Therefore, in this variation we provided participants with English probes that required them to think about the meanings of the gestures they saw. In order to avoid participants relying directly on matches between word order and English syntax, both passive (Was the ball kicked by the boy?) and active (Did the boy kick the ball?) probes were used.

If a noise model like the Swap model described in (2a) reflects actual memory patterns, we expect participants to have an easier time understanding sentences when the roles of agent and patient are easy to disambiguate. In other words, we expect to see an interaction between word order and animacy (reversibility): participants should have particular difficulty remembering SOV order sentences with two animates compared to non-reversible sentences, while no such difference is predicted for SVO order. Of course, since we are working with English-speaking participants there is a possibility that SVO sentences in general will be easier to remember, but the target interaction would still be interpretable so long as participants are not at or near ceiling on the SVO trials. On the other hand, if comprehension of word order does *not* reflect this model, we might see main effects of both word order (SVO easier) and animacy (with reversible/animate-patient sentences being more error-prone), but no interaction between the two.

## *METHODS*

**Participants** 105 English-speaking adults participated on Amazon’s Mechanical Turk.

Participants were screened to be located in the United States and self-reporting English as their first language (an additional 6 participants were excluded who did not meet these criteria.) No other demographic information about participants was collected. The task took approximately 60 minutes and participants were paid \$5 for the study.

**Stimuli** To train participants on the gesture lexicon, we created short movies of the 20 words that were used in the gesture sequences they would see ("Girl", "Boy", "Old Lady", "Fireman", "Heart", "Car", "Ball", "Star", "Elbow", "Kick", "Kiss", "Throw", "Poke", "Fall", "Tumble", "Move", "Jump", "Push", "Lift", "Rub"). Each gesture was at least partially iconic (e.g. the sign for THROW was two arms moving in a throwing motion) and chosen with two criteria: (1) the gestures were similar or identical to gestures used by participants in our gesture production studies and (2) they were distinct from one another.

We also recorded 72 stimulus sets, each consisting of variations on a single event, e.g. “The boy kissed the grandmother.” All events were drawn from the set of nouns and verbs described above. Of these event sets, 24 were transitives with inanimate objects, 24 were transitives with animate objects, and 12 each were intransitives with animate and inanimate subjects.

For each event set, we created target movies in both the SVO and SOV word orders (SV only for intransitive controls) with a single actor pantomiming the entire gesture sequence, pausing briefly between each gesture.

Each movie was paired with a sentence (described in the procedure below.) The 72 sentence items for each participant were randomly selected such that each person saw an even number of animate- and inanimate-patient events in the SVO, SOV, and SV conditions. Probes were constructed for each sentence using passive prompts like “Was the grandmother kissed by the boy?” and active prompts like “Did the boy kiss the grandmother?” For the critical transitive sentences, 3/4 of the sentences were passive, while for the smaller set of intransitive filler trials all prompts were active. This was done so that on the majority of critical trials the probe would

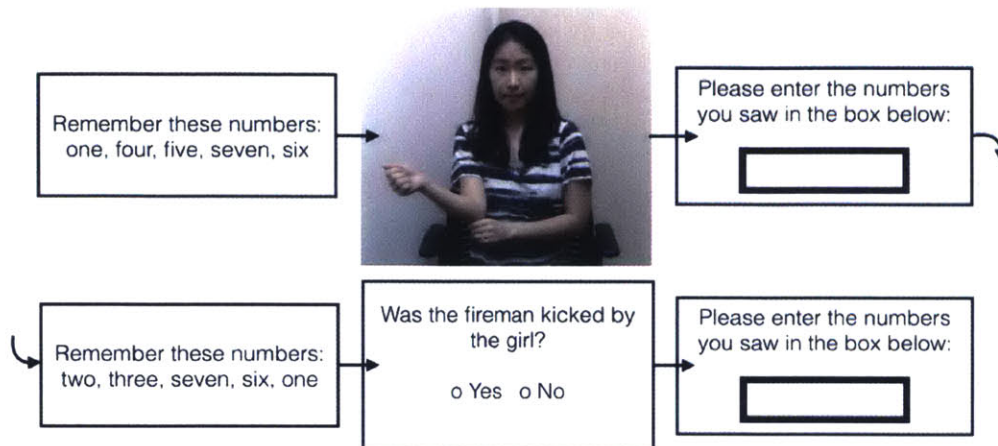


Figure 1: Trial structure for Experiment 1

not match the original gesture sequence for surface order, but participants would not be able to anticipate the kind of question they would be asked.

Prompts were either a correct (1/3) probe, or one of two types of incorrect probe (1/3 Distractor, e.g. “Did the boy kiss the fireman” and 1/3 Reversal, e.g. “Did the grandmother kiss the boy”). Distractors sentences randomly replaced either the subject, verb, or object of the sentence. Reversal probes were used for both animate- and inanimate-object transitive sentences; this led to semantically unlikely prompts like “Was the boy kissed by the ball?”. However, pilot testing revealed that participants did sometimes answer positively to this question in the context of the memory task; this is in line with errors that participants make in other contexts with this kind of passive sentence (Gibson, Bergen, et al., 2013).

All stimuli were presented to participants online using Python and the EconWillow package (Weel, 2008). Code for running the experiment, as well as all stimulus videos, can be found at <http://github.com/mekline/Gesture-Case-marking/>.

**Procedure** After confirming their consent and ability to see videos in the online testing environment, participants were instructed to learn the set of 8 noun gestures that they would be seeing throughout the experiment. They first saw each noun gesture movie paired with the target word and a target picture (this was done to make the gesture referents clearer; e.g. establishing that the GLASSES gesture referred to a grandmother who wore glasses). Then they saw a series

of forced-choice trials with a single video of a gesture and two possible word labels. If they did not choose correctly on all eight trials, participants repeated the items until they had identified the meaning of all the gestures. This train- and test- structure was then repeated for the set of 12 verb gestures.

Once they had learned the gesture lexicon, participants were instructed in the main memory task, which they were told was a test of their multitasking ability. The task consisted of two types of nested memory questions. For the critical Gesture questions, participants saw a movie of a two- or three-gesture sequence, and then later answered a prompt like “Was the grandmother kissed by the boy”. For the flanking Number questions, participants saw a string of five numbers spelled out in words (e.g. FOUR FIVE NINE ONE TWO), and then later typed the number sequence into a text box (“45912”). Each test trial was nested as shown in Figure 1.

Pilot testing revealed that this structure resulted in significant interference on the Gesture questions without reducing responses to chance. After each trial, participants were given feedback on how many numbers they had correctly remembered, and whether or not they had answered the question correctly; this ensured that participants paid attention to both parts of the task. (Comments given at the end of the task indicated that many people believed that the number-memory task was the main point of the study.) At the halfway point, participants took a shorter version of the lexicon quiz to ensure that they still remembered the symbols.

## *RESULTS*

The majority of participants learned the gesture lexicon quickly, successfully identifying all gestures on the first try. For the nouns, the minimum trials to pass training was 8 (a correct answer for each item on the first try), and mean time to pass was 8.2 trials; for verbs, the mean time to pass was 12.4 trials (minimum 12 trials).

Participants attended to both memory tasks, answering both kinds of prompts relatively successfully. For the number task, participants successfully recalled a mean of 3.7/5 items per number list (standard deviation = 1.7 items). For the critical gesture memory task, overall performance was above chance at 75% correct (binomial test,  $p < 0.001$ ).

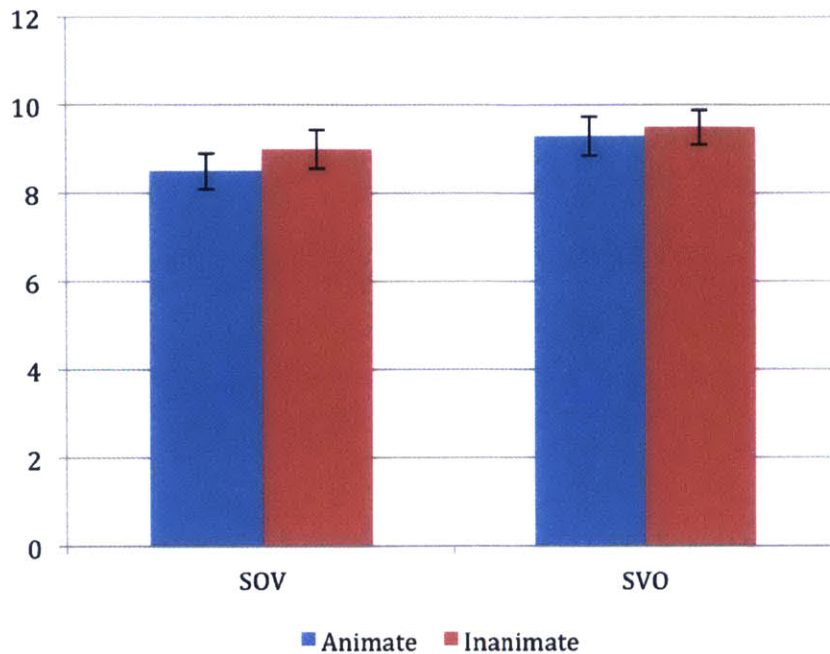


Figure 2: Results of Experiment 1 (Comprehension). We found main effects of sentence order and reversibility (animate/inanimate object), but no interactions. Error bars represent 95% bootstrapped confidence intervals.

The critical analyses focused on differences in recall for gesture items in the two sentence orders (SVO, SOV) and reversibility/animacy conditions (Animate object, Inanimate object.) Using a mixed-effects logistic regression<sup>3</sup>, significant main effects of both Sentence Order ( $X^2=20.9$ ,  $df=1$ ,  $p<0.001$ ) and Sentence Type ( $X^2=4.54$ ,  $df=1$ ,  $p<0.05$ ) were found.

Both of these effects were in the predicted directions: participants successfully recalled more gesture sequences when the order was SVO, or when the sequence had an inanimate patient.

---

<sup>3</sup> A model with full random slopes and intercepts for both participants and stimuli (Barr, Levy, Scheepers, & Tily, 2013) was attempted but did not converge; random slopes were removed until the model converged and all comparisons retained this random effects structure. The final model (in R) was `lmer(wasCorrect ~ sentType*sentOrder + (1|stimNo) + (1|Paycode), data=alldata, family="binomial")`

However, critically no interaction was detected ( $X^2=0.74$ ,  $df=1$ ,  $p=0.39$ ). These results are graphed in Figure 2.

Because the dataset included variations in prompt type (True, False-Distractor Item, False-Reversed) and sentence structure (Passive, Active), we also examined these subsets of the data for additional effects. In all cases the results were qualitatively similar to the overall data, with the exception that no effect of word order was found for the False-Distractor items.

## DISCUSSION

Contrary to the predictions of a noisy channel account, we did not find a selective difficulty in understanding reversible gesture sequences presented in SOV order. We found slight advantages for SVO order (possibly a native language effect) and for non-reversible sequences in general, but no interaction between these factors. These results are problematic for a basic Swap noise model because they show that, at least in this kind of comprehension task, participants do not appear to be making errors of the types predicted (i.e. selectively forgetting/not comprehending one of the nouns, or swapping items that are closer together more than ones that are farther apart.) However, it is important to note that we do not test for these specific comprehension effects directly; rather it is the case that in this comprehension & memory task participants do not spontaneously show the predicted memory errors.

There are a few possibilities for interpreting the lack of the predicted order/animacy interaction. First, as just described the experiment may not adequately model the types of noise (broadly construed) experienced in language or gesture comprehension – the task contains a number of elements, such as the interference task, which don't have clear analogs in comprehension. Second, if the interpretation of gesture production as noisy-channel communication is correct, the *producers'* noise model might not match the actual factors that hinder *comprehension*. Third, it might be the case that noisy-channel frameworks more generally do not explain gesture ordering in production; this data cannot answer this broader question.

## EXPERIMENT 2: PRODUCTION WITH CASE MARKING

In order to clarify the relevance of gesture paradigms for evaluating noisy-channel explanations of human communication and cross-linguistic typology, we need to determine whether the word order effects in gesturing tasks arise primarily from attempts to avoid ambiguity (i.e., from producers operating in a noisy-channel framework GP2013), or from constraints on the production of gestures in particular (i.e. avoiding role-conflict, H2013). As discussed in the introduction, these two views make different predictions about the kinds of task changes that should produce changes in the word orders that participants use. If the SOV/SVO animacy pattern arises from the pressures of noisy-channel communication, word orders should be relatively *insensitive* to changes in the physical gestures used, but sensitive to changes that affect the ambiguity of the signal. If this pattern arises from gesture-specific production pressures, the reverse is true: the particular gestures used to communicate the elements of an event may have a significant impact on the orders used.

To test these predictions, we conducted a study that was designed to modify *only* the information carried by the gestures (i.e., the robustness of the signals to noise.) Despite our attempt, however, analysis of the dataset revealed a potential impact of *both* ambiguity and role-conflict (presence/absence of body-based signs). Analysis of this data and of some of the potential sources for this confound yield important conclusions for how both past and future studies of word order using gesture paradigms should be interpreted.

Experiment 2 aimed to address a puzzle for noisy-channel explanations of word order typology: if the SVO word order is more robust to noise, why aren't all languages verb-medial? GP2013 propose an answer to this question: namely, that the robustness of an SOV-ordered language may increase if the language also possesses a system for marking the event participants for their thematic roles, such as case marking or agreement. Across the world's languages, SOV-dominant languages are much more likely than SVO languages to also use case marking (Croft, 2002; Greenberg, 1963). Critically, the production-constraints account by H2013 does not make predictions with respect to case marking, so long as the signs used continue to involve body-based gestures for animate characters and verbs. In that case, role-conflict between animate patients and verbs (signed from the perspective of the agent) is predicted to yield SVO orders for



reversible events. In contrast, a noisy-channel model would predict that the addition of case marking could make SOV gesture sequences more robust to noise; lowering the pressure to switch away from this plausibly default order.

In the standard gesturing paradigm, some participants spontaneously use spatial cues to mark thematic roles (Hall et al., 2013; Gibson et al., 2013). Gibson et al. found that spontaneous spatial case marking had a strong effect on the gesture ordering: for reversible (animate-patient) events, 64% of the spatially-marked trials used SOV order (cf. 14% of the non-spatially-marked trials). In other words, the use of spatial case marking made participants more likely to retain the default SOV order for reversible events. However, because this evidence comes from spontaneous differences in gesturing, its value is somewhat limited. There might have been other differences between participants who chose to use case-marked gestures and those who didn't, or between items that tended to elicit spatially distinguished gestures and ones that didn't. Thus, the present experiment was designed to experimentally manipulate the use of case marking in the standard gesture task.

The current experiment used the same procedure as in Gibson et al. (2013), but included a component where participants were explicitly instructed to use spatial case marking: they were told to use different hands when signing the event participants in each scene. We analyzed the impact of these instructions on both cue use (i.e. using spatial cues to distinguish characters) and gesture form (use of body-based signs). We predicted that when participants case marked, they would continue to use body-based signs for animate characters, either making a separate gesture with one hand for each character (WAVE LEFT HAND + GLASSES (both hands) = Grandma) or modifying body-based gestures to use only one hand (With left hand: LEFT EYE GLASSES, RIGHT EYE GLASSES = Grandma). However, many participants chose to modify their gestures more extensively –for instance using one arm as a ‘body’ for the character (e.g. establishing LEFT FIST as the first character, then signing HELMET near the first and ROLLER BLADES near the elbow to indicate the boy) or making an inverted V with the first two fingers of one hand that could ‘walk’ around the table to represent human characters.

The critical analyses for this experiment therefore focus on when participants use SOV vs. SVO order, particularly for reversible sentences. These analyses will allow us to determine whether

these changes are due to noisy-channel adaptations, production constraints, or both. In particular, if noisy-channel pressures drive the switch to SVO, then SOV order should be more likely for reversible sentences on those trials where case marking is used; if production constraints drive the shift, SOV order should be more likely when body-based role conflict is *not* present.

#### METHOD

**Participants** 36 native English speakers from the Boston area participated for payment (average age 33 years, 20 female). 13 additional participants were excluded because they did not understand the gesture instructions/did not produce interpretable gesture sequences (7), because they did not complete the entire experiment (2), or because of experimenter error in administering the task (4).

**Materials** The stimuli consisted of 24 animated vignettes shown on a computer screen (created by K. Brink, Gibson et al., 2013). The full list of vignettes is shown in Table 1, and stills from example transitive and intransitive vignettes are shown in Figure 3. Sixteen vignettes depicted

Intransitive events	Transitive events (inanimate patients)	Transitive events (animate patients)
Old lady jumps	Boy kicks ball	Fireman kicks girl
Star jumps	Girl pushes car	Fireman pushes boy
Boy rolls	Old lady elbows ball	Girl elbows old lady
Ball rolls	Girl rubs heart	Old lady rubs fireman
Fireman falls	Fireman throws star	Girl throws old lady
Heart falls	Old lady kisses star	Girl kisses boy
Girl tumbles	Boy pokes heart	Old lady pokes fireman
Car tumbles	Fireman lifts car	Boy lifts girl

Table 1. A full list of the event vignettes used in the verbal/gesture tasks.

transitive events (e.g., *kick, lift*; a total of 8 unique verbs), and eight depicted intransitive events (e.g., *tumble, roll*; 4 unique verbs). Of the transitive events, half involved humans as both the agent and the patient, and were thus potentially reversible (e.g., *boy kicks girl* can be reversed to *girl kicks boy*). The other half had human agents and inanimate patients, such that reversals were not plausible (e.g., *boy kicks ball* cannot be plausibly reversed to *\*ball kicks boy*). In total, there were four different animate characters (boy, girl, fireman, old lady) and four inanimate objects (star, heart, car, ball). Two random orderings of these 24 stimuli were created for the experiment, with an additional two vignettes included as warm-up trials.

**Procedure.** The experiment consisted of three components: a verbal task and two gesture tasks. The exact text of the instructions given before each task is shown in Table 2. In the verbal task, participants were simply asked to verbally describe each of the 26 vignettes (2 warm-up + 24 target) in a simple sentence. In the first gesture task, they watched the vignettes again and were asked to act out each event using only gestures and no speech. No specific instructions were given about the types of gestures that participants should use, and the experimenter did not demonstrate any sample gestures or refer to parts of speech or thematic roles such as subject/verb/object or agent/action/patient. In the final task, participants watched the vignettes again, and were instructed to include spatial information in their gestures. Again, the instructions

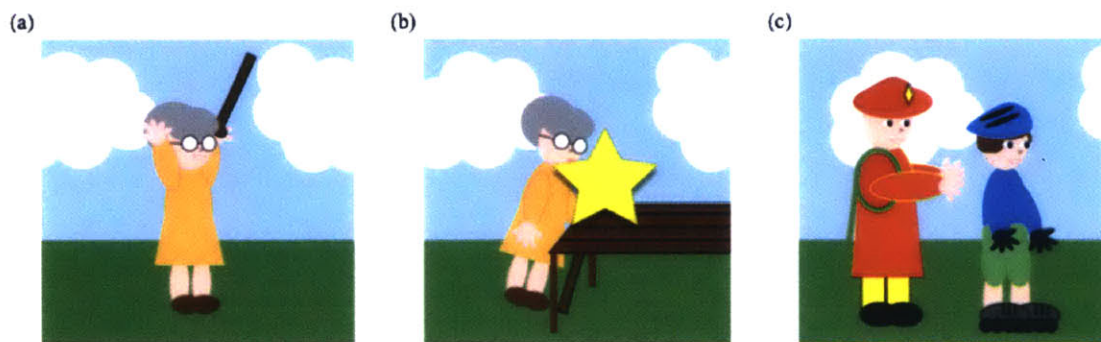


Figure 3. Three examples of stills from the animated event vignettes used in the experiment. (a) *Old lady jumps* – Intransitive (b) *Old lady kisses star* – Transitive, inanimate patient (c) *Fireman pushes boy* – Transitive, animate patient

did not include any references to specific parts of speech or thematic roles. Instead, participants were told that they should use one hand when they need to gesture about a “person or thing”, and use the other hand whenever they needed to gesture “another person or thing”.

**Coding.** All responses were filmed and independently coded by two of a group of four trained coders, including at least one coder blind to condition (for blind coding, the videos were clipped into sequences of four trials from either the first or second gesture task, and then interleaved in a random order). Responses to each scene were coded to reflect the order of gestures referring to the agent, patient, and action (using S, O, V terminology throughout). Thus, a series of consecutive gestures referring to one concept were coded as a single instance of S, O or V (e.g., a gesture sequence *GLASSES-CANE-BUN* all referring to the old lady would be coded as a single S or O gesture). Participants occasionally paused or re-started their gesture sequences, in which

General Instructions:

*You are going to watch a series of 24 short videos on the screen. Your job is to explain each scene to a person who can't see the screen so that they would know what's going on.*

Verbal Description Task:

*We're first going to start with the verbal description part. After each video, tell me what happened in one sentence.*

Gesture Task 1:

*Now you are going to watch the same videos over again. This time, describe what happened using your hands. This means you are not allowed to say anything out loud, just like playing charades. Please make sure you are detailed enough so that someone who can't see the video would know what happened. Try not to talk or mouth along with your gestures.*

Gesture Task 2:

*In this part, you're going to watch the same slides over again, but this time I would like you to include location information in your gestures. So, whenever you need to mention an object or person, mark its location using one hand. And to mark another object or person, use the other hand. Again, please make sure you are detailed enough so that someone who can't see the video would know what happened, and try not to talk or mouth along with your gestures.*

Table 2. Instructions given to participants before each phase of the description task

case their final continuous gesture sequence was analyzed (re-starts were coded as any instance where the participants put their hands down and stopped gesturing for more than two seconds, or waved their hands to signal a 're-do').

1148 gesture sequences whose content could be determined (plus 4 illegible sequences) were produced by the 36 participants. On the majority (833 items, 73%) of trials, participants' responses consisted of exactly one expression of S, O, and V in some order. However, in some cases participants either omitted a referent (e.g. VO), included multiple repetitions (e.g., SOSV) or produced simultaneous gestures (e.g., simultaneous SO followed by V). In order to code these remaining items, we asked whether the response a participant produced could be sensibly classified to evaluate the central hypothesis of the experiment, namely whether people preferentially separate subjects and objects with a verb when both are animate. Thus, wherever possible the remaining utterances were classified as either Verb Medial (VM, e.g., SVO/OVS) or Verb Non-Medial (VNM, e.g., OSV, VOS, etc.). To do this, we restricted responses to the final three gestures of the final gesture sequence produced on each trial, and checked whether this response included all three referent types in a code-able order: for instance, simultaneous S and O followed by V would be coded as Verb Non-Medial, whereas S followed by simultaneous O and V would be left unclassified. Responses consisting of only one or two gestures were left unclassified since the position of the two nouns relative to a verb could not be defined<sup>4</sup>. Overall, this process allowed for the inclusion of an additional 124 items for a total of 957 items in the analysis; none of the effects reported below are significantly affected by the inclusion or exclusion of these 124 items.

---

<sup>4</sup> This final-three-gesture coding was designed to capture the relevant information about each gesture sequence (position of V relative to S/O) in a straightforward way. 133 gesture sequences were produced that were longer than 3 gestures (e.g. SOSOV), of which 97 could be classified under this metric as Verb-Medial or Verb-Nonmedial. Because these longer sequences typically consisted of repetitions of S and O at the beginning of the sequence, the truncation usually either reflected the overall order (e.g. SOSOV -> SOV) or resulted in the 'unclassified' coding for marginal SVO sequences – those where one of either S or O appears *both* before and after a verb, but the other has a unique ordering (e.g. SVOV -> VOV). Of the 97 successfully classified items, only three tokens appear potentially questionable under our classification scheme (two instances of SOVSVO classified as Verb-Medial; one instance of OVOSV classified as Verb-Nonmedial).

Inter-coder reliability was calculated over both raw gesture codings and over final classifications (VM, VNM, Unclassified). On the raw codings, coders agreed on 78% of trials (Cohen's kappa = 0.72), and on the final classifications coders agreed on 88% of trials (Cohen's kappa = 0.80). All disagreements were resolved by discussion with a third member of the coder group who remained blind to condition.

All responses were also coded to reflect whether the participant used spatial information to distinguish the subject and object in a sentence. The second gesture task included specific instructions to use a "right hand/left hand" marking system, but participants also spontaneously developed alternative spatial systems during both gesture tasks, consistent with prior reports (e.g., Hall et al., 2013; Gibson et al., 2013). Responses were coded as including spatial/case marking information if participants added a specific marker (e.g., 1 finger or 2 fingers following a noun sign) or produced the sign for an agent or patient (at least one) in a specific location away from their neutral signing posture, by gesturing to a new location in front of them or by shifting their torso (following possibilities for spatial distinguishing of referents in American Sign Language and International Sign pidgin, Lillo-Martin & Klima, 1990; McKee & Napier, 2002). Thus, we did not code incidental differences in location due to the production of particular signs (e.g., HELMET, which was typically gestured in an upper part of the signing space vs. ROLLERSKATES, which were typically gestured in a lower part of the space) as 'spatial information'. Because this variety of case marking strategies was discovered during the blind-coding process described above, a new coder who was blind to the hypothesis and condition coded 10% of the data (4 participants) for spatial/case marking information. Agreement was very high (92%, Cohen's kappa = 0.84) and remaining disagreements were resolved by discussion among blind-to-condition coders.

Finally, each gesture 'word' (S,O, or V sequence) in each response sequence was coded as either body-based or not body-based. For nouns, gestures were coded as body-based if the distinctive features of a person (e.g. accessories, hairstyles, postures) were located on the signer's own body. If gestures for people were placed in the signing space in front of the participant, or indicated in a way that didn't place a character's feature in the analogous place on the participant's own body, the gesture was coded as not body based. For instance, a sign for

HELMET (indicating the boy wearing a bicycle helmet) might be produced by gesturing a helmet on your head (body-based), by showing the shape of a helmet in front of you (not body-based), or using your arm to indicate the boy and placing the other hand over your fist as the helmet (not body-based). 30% of responses were coded for the body-based judgments by two coders, one blind to condition (as described above); inter-coder agreement was very high (96%, Cohen's kappa = 0.865).

## RESULTS

### 1. Word Order

We first evaluated whether the case marking instructions led to a general shift in the distribution of word orders produced by participants. Because this study focuses on evaluating the main prediction of the noisy-channel model (the SOV/SVO animacy pattern) rather than on other word orders that have been observed in gesture tasks, we focus on the classification of gesture orders as either Verb Medial (VM, i.e. those which are informative because they order one participant before the verb and one after, like SVO) or Verb Non-Medial (VNM, those which place both

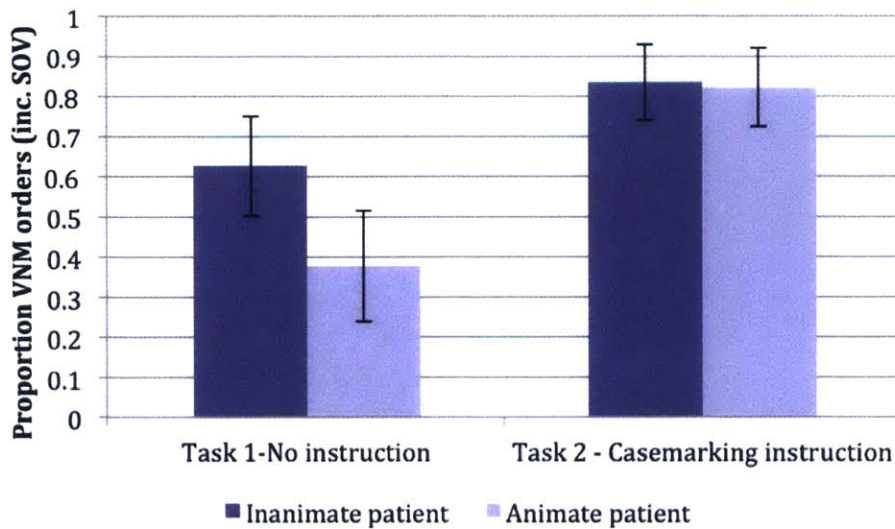


Figure 4. Proportion of gesture sequences using SOV/VNM order in the first (free gesture) and second (spatial/case marking instructions) gesture tasks. Error bars represent bootstrapped 95% confidence intervals around the means of each condition.

participants on the same side of the verb, like SOV.) The classification of gesture orders is described above, but the majority of items in these categories consist of SOV and SVO orders, and results do not change if only these orders are evaluated.

The word order results are summarized in Figure 4. During the first gesture task, results for both inanimate and animate patients closely paralleled the findings of previous studies of word order in gesture (Goldin-Meadow et al., 2008, Gibson et al., 2013). This difference in VNM gestures (63% versus 38%) was statistically significant in a two-tailed mixed-effects logistic regression that included random participant slopes and item intercepts,  $X^2 = 18.979$ ,  $df = 6$ ,  $p < .001$  (Gelman & Hill, 2007).<sup>5</sup>

In the second gesture task, after being instructed to use spatial information to mark the entities in the events, participants tended to produce VNM sequences regardless of the patient's animacy. The difference in verb-final gestures for animate and inanimate patients was not statistically significant for this task ( $X^2 = 0.07$ ,  $df=6$ ,  $p = .79$ ).

Analyzing the two gesture tasks together, we found a significant task by animacy interaction ( $X^2 = 16.91$ ,  $df = 6$ ,  $p < .001$ ), as well as main effects of animacy ( $X^2 = 18.22$ ,  $df = 5$ ,  $p < .001$ ) and gesture task ( $X^2 = 184.1$ ,  $df = 5$ ,  $p < .001$ ), indicating that participants were more likely to use SOV/VNM order in the second task.

## **2. Gesture features of interest - Spatial marking and body-based signs**

---

<sup>5</sup> Animacy was varied within subjects, but not within items. The model with maximal random effects can thus be expressed in R as `lmer(WordOrder ~ Object.Type + (Object.Type|Subject) + (1|Sentence), family="binomial")`. This analysis is used for all remaining effects reported in the experiment: the model with full random participant slopes and intercept and item intercepts is reported unless it fails to converge, in which case random effect factors are removed until convergence is reached. Full details of all analyses are available at [github.com/mekline/Gesture-Case-marking](https://github.com/mekline/Gesture-Case-marking)



We next evaluated the two possible mechanisms for word order, namely (1) whether participants produced spatially coded gestures (i.e. case marking) on individual trials and (2) whether they used body-based gestures. As noted above, despite receiving no instructions to mark characters with spatial information in the first gesture task, and explicit instructions to do so in the second, participants in fact varied in their use of spatial information during both tasks. Likewise, participants in previous gesture tasks (e.g. H2013) have been observed to use body-based signs, and we expected that this would be the case in both of our gesture tasks.

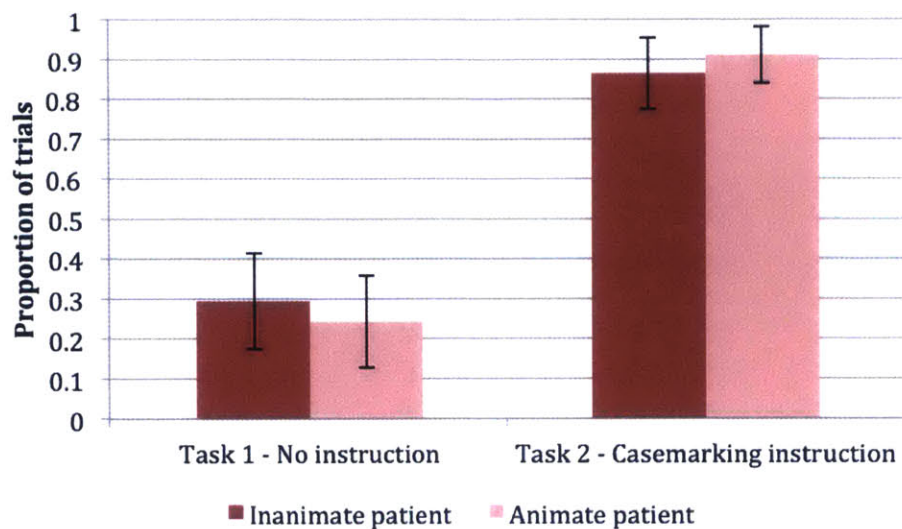


Figure 5: Analysis of spatial cueing: proportion of trials using spatial cues in the first and second gesture tasks.

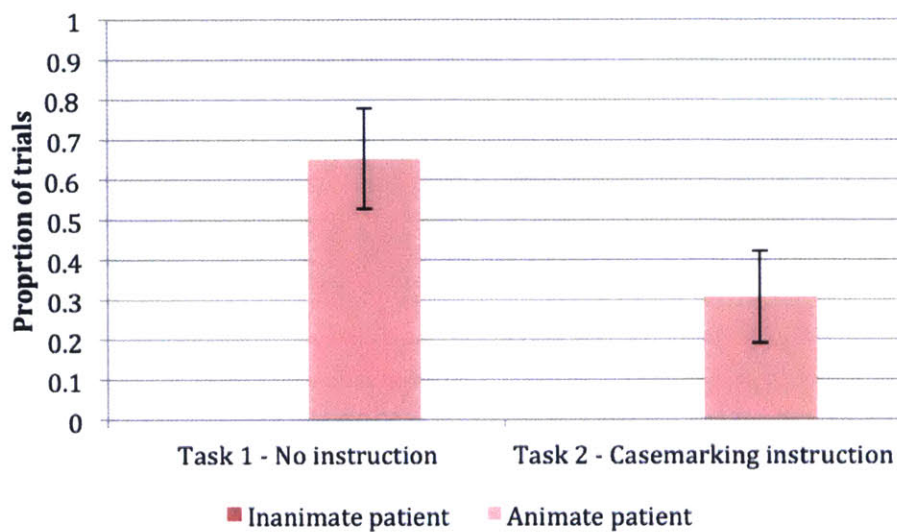


Figure 6: Analysis of role conflict potential: proportion of trials with role conflict (use of body-based gestures) in the first and second gesture tasks. Note that there is no potential for role conflict on inanimate-patient trials.

Our intended ‘case marking’ manipulation was successful: after receiving instructions before the second gesture tasks, participants produced many more gestures with spatial information (Figure 5). Interestingly, participants did not restrict their use of spatial information/case marking to animate-patient trials. There was a very large effect of gesture task ( $X^2 = 43.05$ ,  $df = 7$ ,  $p < .001$ ), such that participants were much more likely to use spatial gestures after they had been instructed to do so. Thus, case marking as a strategy became more common for both animate and inanimate trials in the second gesture task. However, somewhat consistent with the predictions of a noisy-channel model, there was a small but significant interaction between task and animacy ( $X^2 = 6.29$ ,  $df = 8$ ,  $p < .05$ ): after instruction, participants’ use of spatial information increased somewhat more for animate trials in particular.

As discussed above, our experimental manipulation can only provide unique support for the noisy-channel model if case marking does *not* also lead participants to produce non-body-based gestures. H2013 describe the potential for role conflict when both an animate patient and the immediately following verb are gestured with body-based signs, because the perspectives of the patient and the verb conflict for the producer. Our measure of ‘body based’ gesture sequences is therefore the union of patient and verb body-based signs, referred to hereafter as ‘potential for role conflict’.

We found that this potential for role conflict within animate trials declined significantly from the first gesture task to the second. ( $X^2 = 26.8$ ,  $df = 6$ ,  $p < .001$ ). This is summarized in Figure 6; note that since role conflict can only occur when the patient is animate that the values for inanimate patients in both tasks is 0%.

### **3. Relationship between word order and gesture features**

We have established that the patterns of word orders changed between gesture tasks, and also that both gesture features of interest (case marking and potential for role conflict) changed between tasks as well. While we cannot establish causality, we nevertheless attempt to describe the relationship between case marking, role conflict, and word order. To do this, we limit the analyses to the Animate trials because the effects of both gesture features are hypothesized to apply to these kinds of events in particular.

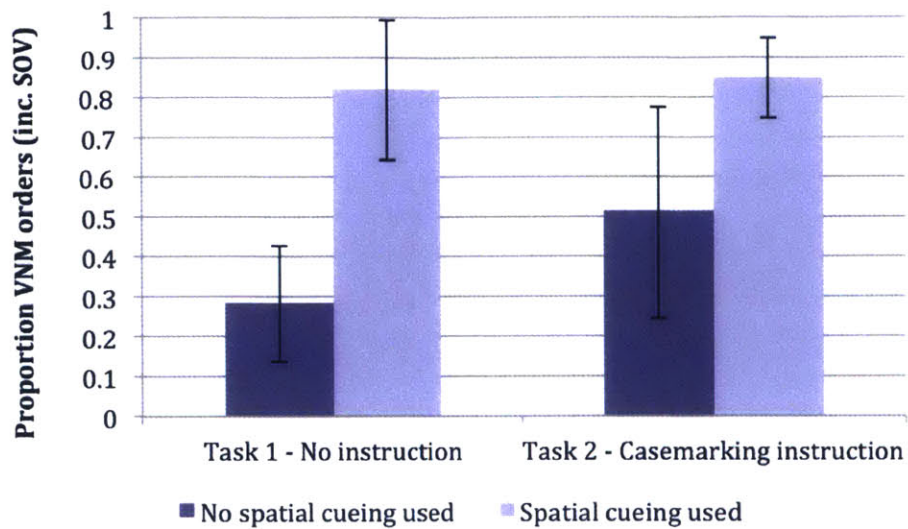


Figure 7. Word order and spatial cueing: Proportion of VNM order on reversible trials when spatial information was or wasn't included.

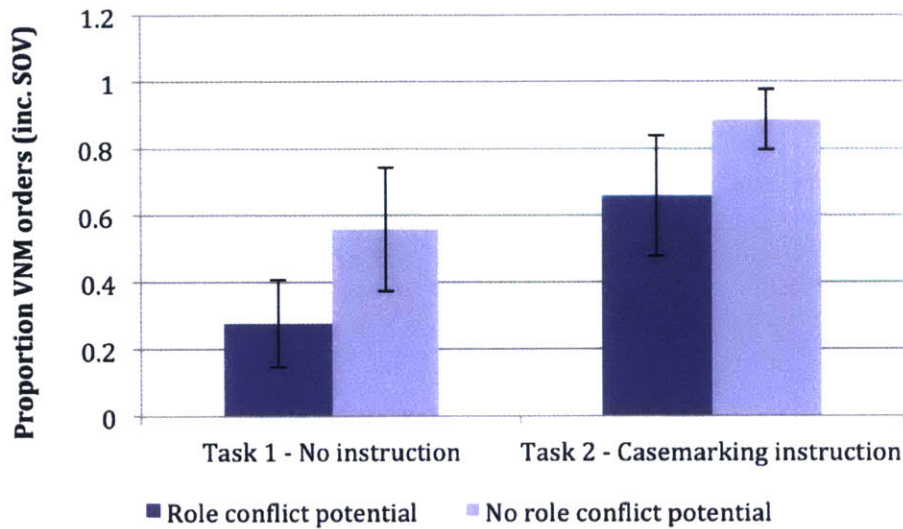


Figure 8: Word order and role conflict: Proportion of VNM order on reversible trials when role conflict potential was present or not present.

The use of spatial information (case marking) in gesture sequence was closely related to the gesture orders that participants used: in both experiments, participants were more likely to gesture in a VNM order when they also used spatial information; No-instruction task  $X^2=18.21$ ,  $df=6$ ,  $p < 0.001$ ; Case marking instruction task  $X^2= 8.33$ ,  $df=6$ ,  $p < 0.01$ ). This information is summarized in Figure 7. There was no interaction between task and case marking on the gesture order used ( $X^2=0.05$ ,  $df=8$ ,  $p=0.82$ ).

Role conflict was also closely related to gesture order: Participants were more likely to use a VNM order when they *avoided* role conflict.<sup>6</sup> This difference was significant in the No-instruction task ( $X^2= 6.47$ ,  $df=6$ ,  $p < .001$ ) and marginally significant in the Case marking instruction task ( $X^2 = 3.39$ ,  $df=4$ ,  $p=0.07$ ); there was also significant interaction between role conflict and task ( $X^2 = 3.97$ ,  $df= 8$ ,  $p< 0.05$ ). These results are summarized in Figure 8. We plot

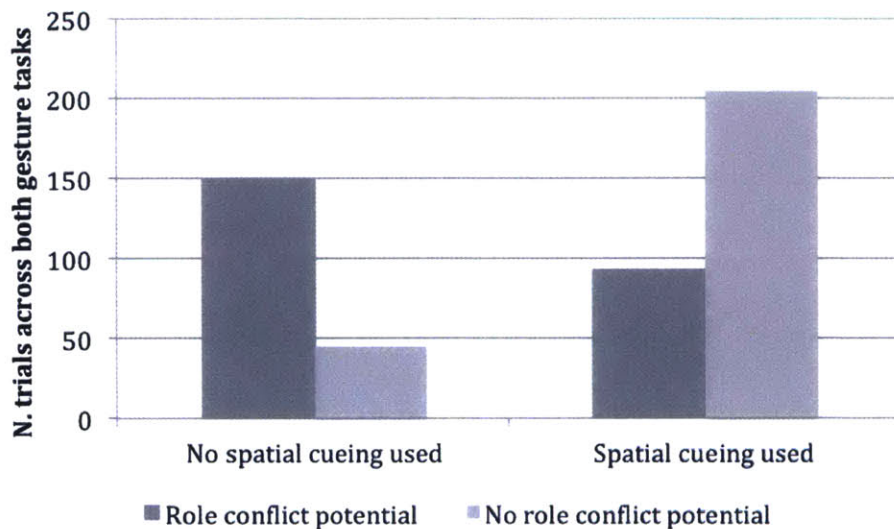


Figure 9: Histogram of gesture trials (across both tasks) with role conflict potential (body-based gestures) and spatial cues.

<sup>6</sup> Under H2013, the classification of the rarer word orders (e.g. not SOV and SVO) is somewhat different than under a noisy channel theory. We use the same classification of gesture orders throughout for clarity and because these other orders are not particularly frequent in our dataset; however re-dividing these items by the “avoid OV” role-conflict metric yields qualitatively similar results for all analyses.

trials with potential for role conflict on the left to facilitate comparison between the two mechanisms (compare to Figure 7): *absence* of role conflict and *presence* of case marking are both predicted to increase SOV orders for animate-patient trials under their respective theories.

#### 4. Contingency between mechanisms

Finally, we explored the contingency between role-conflict and case marking, both features of the outcome gestures that might explain the higher instance of SOV outcomes in animate trials in the second gesture tasks. These features were highly related: role conflict negatively predicted case marking ( $X^2 = 19.36$ ,  $df=6$ ,  $p < .001$ ). In other words, when people produced gestures with role conflict (i.e. body based gestures), they were less likely to use spatial information to differentiate their gestures for event participants, and vice versa. This is summarized in Figure 9.

#### DISCUSSION

In this experiment, we first replicate the critical gesture pattern (Gesture Task 1): participants shifted away from the default preference for verb non-medial gestures (typically SOV order) to the SVO order (or rarely other verb-medial orders) when describing semantically reversible events (i.e., events with two animate characters, where the thematic roles are potentially ambiguous).

Previously, GP2013 argued that this pattern reflects a tendency to produce gesture sequences that make it easy to recover the correct meaning under noisy conditions. For non-reversible events where patients are inanimate, the action could only occur in one plausible way—a boy can kick a ball, but a ball cannot kick a boy. Thus, if an SOV message were corrupted by Swap-type noise, *BALL BOY KICK* is interpretable on semantic grounds. However, this is not the case for reversible events (a girl can kick a boy and a boy can kick a girl). Using SVO order is one strategy that can improve meaning recoverability – if adjacent elements are swapped (*GIRL KICK BOY* -> *GIRL BOY KICK*), the relative order of the agent and patient will be maintained, which a comprehender can use as a cue to the intended meaning.

Our goal in manipulating case was to tease apart the noisy-channel and role-conflict hypotheses by manipulating the ambiguity of the sentences while leaving the form of participants' gestures alone. However, the instructions in fact lead to significant changes in how participants gestured

animate characters – in addition to using just one hand, they often moved away from body-based signs, e.g. gesturing BOY as an inverted V ‘walking’ on the table rather than as a bicycle helmet on their own head. This led to an opportunity to attempt to evaluate the impact of *both* noisy-channel and role-conflict hypotheses.

The two critical features of gesture (using spatial information and *avoiding* body-based signs) occurred during both gesture tasks, increasing following the instructions given for Gesture task 2. The use of these gesture features also varied within sessions and within participants, allowing us to evaluate whether the use of a particular gesture order (SOV or SVO) patterned with these features. Critically, we found that these two gesture features were correlated with one another. When people produced gestures with role conflict (i.e. body based gestures), they were *less* likely to use spatial information to differentiate their gestures for event participants, and overall more likely to produce SVO order for reversible events. When they used spatial differentiation, they also moved away from the body-based gestures that could produce role conflict, and they tended to produce SOV orders for animate-patient events. Thus, these data are ambiguous between the explanations offered by GP2013 and H2013; because of the concomitant changes in gesture form, features of gestures that provide an ad-hoc ‘case-marking’ system do not necessarily support the idea that producers are trading off between strategies to make their gestures less ambiguous.

Another feature of the data that does not straightforwardly support a noisy-channel interpretation is the fact that participants used spatial information at the same rates for animate and inanimate gestures: a more strictly rational strategy that took the extra cost of producing ‘case-marking’ information only when it was necessary would produce it only on animate trials. In many natural case-marking languages this is the case: non-default case may only be marked when patients are animate (cf. Malchukov, 2008). A proponent of the noisy-channel interpretation of this data might argue that as in many of the world’s languages, case-marking can also be adopted as a general strategy to support a less informative word order. However, this means that even this additional exploratory analysis does not support the noisy channel interpretation over the production constraint (role-conflict) account.

Of course, it is possible that *both* proposed mechanisms were acting in concert to produce the observed patterns. That is, it might both be the case that gesturing tasks have particular constraints that affect ordering, and that more general communicative constraints following a noisy-channel model influence ordering as well. Alternately, either one of these explanations might be producing the effect alone. Since creating a manipulation which affected only ambiguity and not gesture form has proven to be difficult, we therefore next take the opposite approach.

### EXPERIMENT 3: PRODUCTION WITHOUT BODY-BASED GESTURES

To see whether a noisy-channel account can account for gesture patterns alone, we therefore tested whether production constraints *alone* could produce changes in word order. In order to preserve as much of the standard gesture task as possible, we use a physical constraint rather than altering the instructions given to participants: a curtain that prevented participants from using any body-based gestures. Under a noisy-channel communication hypothesis (GP2013), this change is *not* predicted to affect word order: the signs that participants create will still be able to express concepts like BOY and THROW, and the same communicative pressures against having two animates next to each other (assuming a Swap noise channel) would apply. If anything, we might expect a *larger* effect of SVO in animate sentences, if the symbols chosen for e.g. BOY and GIRL are more difficult to distinguish. On the other hand, under the role-conflict theory, making all gestures away from the body (in the signing space in front of the participant) should reduce the production pressure that H2013 proposed leads to SVO orders for reversible sentences. Instead, this constraint would allow participants to produce events of all types in the more canonical or cognitively basic SOV order.

As in Experiment 2, we coded for the critical gesture features that are hypothesized to mediate gesture order. We intended for the manipulation to make body-based signs impossible, and this was the case – participants did not attempt body-based signs. We also coded each trial for the presence or absence of spatial information, in order to test for relationships between case marking and word order.



## *METHODS*

**Participants.** 17 native English speakers from the Boston area participated for payment (average age  $n$  years,  $n$  female). 5 additional participants were excluded because they did not understand the gesture instructions/did not produce interpretable gesture sequences (2), because they did not complete the entire experiment (1), or because of experimenter error in administering the task (2).

**Materials.** The stimuli were identical to those used in Experiment 2. In addition, an approximately 9in x 14in curtain frame was created from PVC pipe and felt, with two slots for participants to insert their arms through.

**Procedure.** The procedure for Experiment 3 was identical to the first half of Experiment 2. The instructions were the same as in the basic gesture task, with one addition: participants were asked to place their hands through the curtain, and told that only gestures made with their hands and arms were being filmed. Coding of gesture identity and order was conducted in the same way as in Experiment 2, and gesture orders were classified according to the same Verb-Medial/Verb-Non-Medial scheme.

By hypothesis and as anecdotally confirmed during testing, body-based signs could not be produced by participants in this experiment. Spatial cueing was possible (while body shifts could not be used, participants could potentially use one-handed gestures or other hand signs to distinguish patients from agents) and so was coded following the same procedures as Experiment 2.

## *RESULTS*

### **1. Word Order**

The main analysis for this experiment concerned whether the distribution of word orders produced by participants matched or diverged from the well-established pattern for animate- and inanimate-patient events reported in previous studies (Gibson et al., 2013; Hall et al. 2013; Meir et al. 2010). Again, we classified gesture orders as either Verb Medial (VM, i.e. those which are informative because they order one character before the verb and one after, like SVO) or Verb

Non-Medial (VNM, those which place both character on the same side of the verb, like SOV) in order to evaluate the claims of the noisy-channel model. However, the majority (69%) of the gesture orders were exactly SOV or SVO), and all analyses reported below are qualitatively similar if only those gesture responses are included.

The word order results for Experiment 3 are shown in Figure 10. Unlike previous findings, when participants were required to execute their gestures through the curtain, we found *no difference* between the prevalence of VNM orders for animate-patient and inanimate-patient items ( $X^2 = 0.37$ ,  $df = 6$ ,  $p = 0.84$ ). In fact, there was a numerical difference in the opposite direction than predicted by noisy-channel models: participants were slightly more likely to use SOV and related orders when they gestured for events with an *animate* patient.

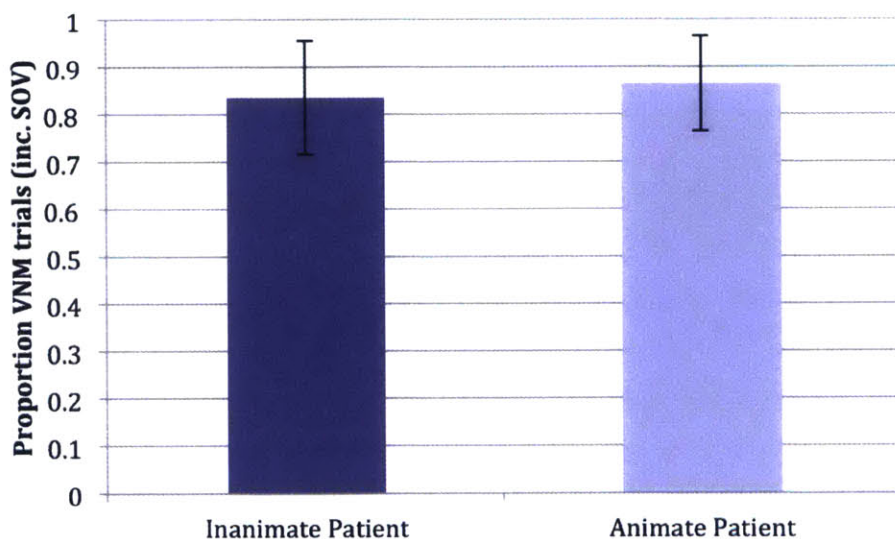


Figure 10: Proportion of Verb Non-Medial (e.g. SOV) orders for animate- and inanimate-patient events when body-based gesturing was prevented (Experiment 3).

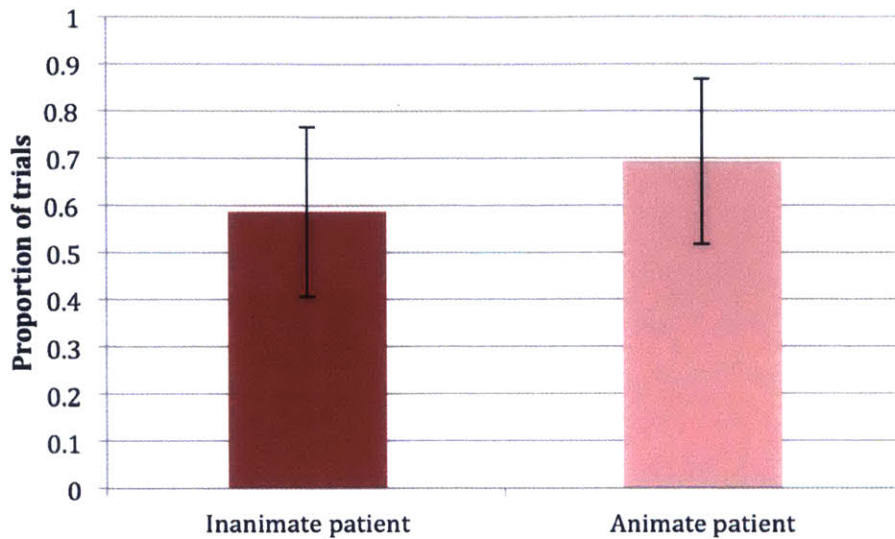


Figure 11: Analysis of spatial cueing (Experiment 3): Proportion of trials using spatial cues

Next, in parallel with Experiment 2, we asked whether the use of spatial cues to distinguish the subject and object (case marking) was related to the word orders participants used. The first finding is that participants used case marking far more often than in the standard gesture task: 61% of trials included case marking; comparing this pattern to the case marking data from the standard gesture task in Experiment 2, there was a significant main effect of experiment ( $X^2 = 11.3$ ,  $df = 4$ ,  $p < 0.001$ ). We next asked whether the pattern of case marking in the curtains task differed with reversibility (animacy of the patient); as shown in Figure 11, there was an effect of animacy such that case marking was used more often with reversible events ( $X^2 = 6.49$ ,  $df = 4$ ,  $p < 0.05$ ). However, recall that this effect was not found in either of the tasks in Experiment 2, so this should be interpreted cautiously.

The use of case marking was also related to the word orders that participants produced: within Animate-patient trials (the critical test case for both noisy-channel and role-conflict models). There was a significant effect of case marking on word order (Fisher's exact test,  $p < 0.001$ ): when participants used case marking (as they did for approximately 70% of reversible trials) they *always* produced SOV or other Verb Non-Medial orders. This information is summarized in Figure 12.

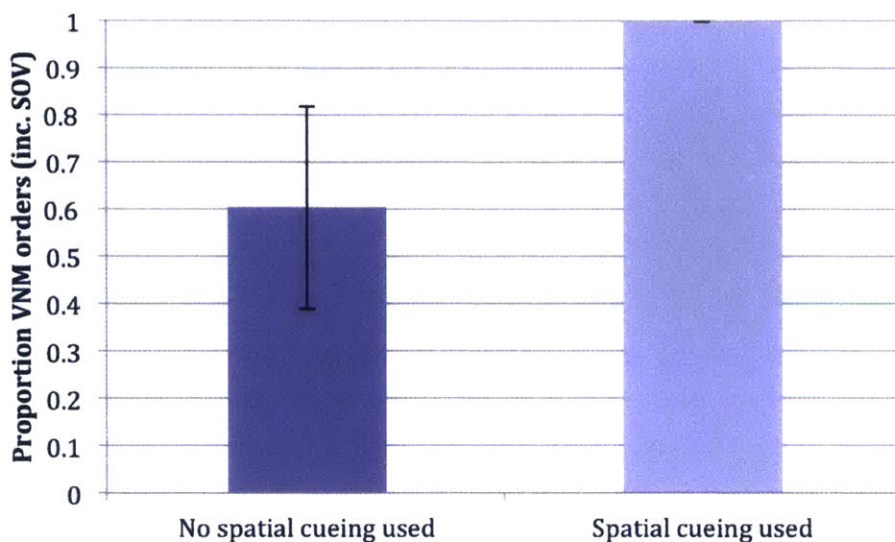


Figure 12: Word order and spatial cueing: Proportion of VNM order on reversible trials when spatial information was or wasn't included.

### DISCUSSION

Contrary to the predictions of the noisy-channel model, and in line with the gesture-specific production constraints proposed by H2013, requiring participants to make their gestures in front of them instead of with their whole bodies produced a significant change in gesture order. Although no other changes to the event stimuli or instructions were made, participants used primarily SOV and related orders for both animate- and inanimate-patient events, if anything trending toward using more SOV orders for animate-patient scenes. Under H2013's model this is transparent: because no body-based signs were used, participants did not experience role conflict when transitioning from one sign to the next, and so had no reason to switch from the cognitively-basic SOV word order to SVO. Under a noisy channel model, this is less easy to explain: if the gestures are treated as holistic representations of the concepts (i.e. the participant really is just communicating concepts like BALL and THROW with their gestures), it is difficult to understand why executing these signs with the hands rather than with the body would eliminate the pressure to reduce noise that otherwise drive the shift to SVO.

However, a closer examination of the gestures produced in this experiment yield an insight parallel to Experiment 2: when gesturing through the curtain, participants often used spatial information, choosing to gesture event participants with just one hand and often alternating between hands to distinguish them. This pattern is in line with the type of spatial marking that GP2013 proposed could serve as an alternative to SVO order to make gestures robust to noise from a simple Swap (or Deletion) noise channel. In addition, while participants usually case marked and usually used SOV orders for reversible trials, the correlation of these two features is in line with a noisy-channel model: when participants did *not* include spatial information, they were more likely to use SVO order.

The cause of these changes is unclear: perhaps participants felt that since their hands were held in relatively fixed positions, they should treat each as a more separate gesturing tool. Or, perhaps the curtain seemed like a source of additional confusion and encouraged participants to make their gestures clearer by adding spatial information – we cannot tell from this dataset which is the case. In short, this experiment makes it clear that with respect to the gesture paradigm, we are in one of two troublesome situations: either gesture order effects are the result of gesture-specific production constraints alone, or they are the product of a noisy-channel communication effect *plus* a series of factors specific to gesture, factors which are hard to eliminate from this naturalistic task and hard to predict when designing experiments.

## GENERAL DISCUSSION

Over the past several years, a robust pattern of effects have emerged concerning the word orders that participants from multiple language groups use when gesturing simple transitive events. When agents are animate and patients are inanimate, people tend to use SOV orders regardless of their native language, but when both agents and patients are animate (reversible events), people tend to switch to SVO order (GP2013; H2013; Meir, Lifshitz, et al., 2010). While the results have been replicated several times, the explanations for these patterns have been in significant dispute. Many of these explanations have relied on the assumption that gesturing by naïve participants reflects modality-independent word order preferences (GP2013; Langus & Nespors, 2010; Meir, Lifshitz, et al., 2010). However, it has also been pointed out that these word orders

may result from the specifics of the gesture task itself, and the particular challenges it creates for producers participating in the task (H2013).

Here, we present three experiments that reveal some challenges for interpreting gesturing orders within the noisy-channel communication framework. In Experiment 1, we found no evidence that reversible gesture sequences in SOV order were especially difficult to remember or understand. To interpret gesture patterns as a noisy-channel phenomena, it is necessary to specify the kind of noise model that a producer might be assuming when they construct their utterances. We consider two simple processes, Deletion and Swap channels (see examples 2a and 2a) that might make SVO order easier to comprehend than SOV. While a noisy-channel model need not involve producers who are *correct* about the challenges that comprehenders face, the fact that comprehenders do not appear to be sensitive to word order in this task (even when the more difficult order also departs from the English speaking participants' native language) is surprising. If the task successfully models important aspects of comprehension, this suggests that either (1) producers do not use a noise model like those we proposed, (2) producers do use such a model, but comprehenders do not actually suffer the predicted kind of noise effects, or (3) noise models are not a central pressure on how producers gesture simple sentences.

To determine which of these explanations are most likely, we then attempted to establish whether or not the (production) gesture task results could be explained by noisy-channel models. In fact, the results of these two experiments reveal a potential effect of the particular communication medium, gesture, on the word orders produced. In the first gesture production experiment (Experiment 2), we added a 'case-marking' manipulation which was designed to affect only the *ambiguity* of the gestures, by encouraging participants to differentiate the agents and patients from each other. Under this manipulation, participants tended to use SOV orders even for animate-patient scenes, as predicted by a noisy-channel model. However, we discovered a second mechanism that might have produced the same effect: the instructions also led participants to produce signs away from their own body, reducing the role-conflict production constraint proposed by H2013. To differentiate between these mechanisms, we next implemented a manipulation (Experiment 3) which was intended to affect only the *physical form* of the gestures. A noisy channel model would predict this to have no effect on the word orders

used, but we found a significant impact: asking participants to produce gestures with just their hands/arms, instead of their whole bodies, entirely eliminated the animacy-dependent SOV/SVO pattern: participants produced mainly SOV order across the board. However, this experiment was also ambiguous: we observed participants using spatial information to distinguish event participants more often than in the standard task. This means that it is possible in principle that the SOV order resulted from this additional source of information marking the identity of agents and patients.

The goal of the present studies was to determine whether studies using the gesture paradigm developed by Goldin-Meadow et al. (2008) can yield evidence on the source of typologies of basic word order in the world's languages. For this to be the case, it has been assumed in previous studies by Gibson et al. and others that the gesture orders produced by participants are a relatively straightforward translation of the order in which they might intend to communicate concepts if they were not influenced by the orders of their native language. However, gesturing is unique in at least one way as a communication task: first, unlike spoken language and to a greater degree than developed, grammaticalized sign languages, the gestures that people produce are usually iconic. Therefore, they sometimes express multiple meaning elements together (e.g. signs for verbs that position the gesturer as the agent, as described by H2013). This iconicity may produce additional constraints on production or communication; in fact, it has been proposed that the strong SOV bias in developed, grammaticalized sign languages may be due to the visual iconicity present in manual languages (Cuxac & Sallandre, 2007; Napoli & Sutton-Spence, 2014; Taub, 2001).

In Experiment 2, we hypothesized that introducing an ad-hoc case marking system would improve meaning recoverability, and thus participants would produce more SOV gestures even for reversible events. However, our experiment had a critical pre-requisite for interpretation: that the general form of people's gestures remain roughly the same, and therefore orthogonal to H2013 account of gesture-specific 'role conflict'. While the particular form of gestures produced in these gesture tasks has not previously been formally analyzed, previous studies with both English and non-English speaking populations have reported largely similar gesture repertoires.

However our manipulation produced more extensive changes, eliminating the kinds of gesture features that might present a specific conflict for producers.

These large changes in gesture form created a significant problem for interpretation of the present experiments, but also reveal that results in the standard gesture task do not represent modality-independent or ‘natural’ communication order alone. Together, studies using the gesture task have produced some novel insights into how people communicate (or attempt to communicate) in an ad-hoc system outside their native language. The cross-linguistic consistency in gesture patterns is striking and well replicated, particularly the finding that SOV order is primary or default for default transitive scenes (e.g. an animate agent causing a motion or change of state in an inanimate patient). Converging evidence from existing languages and language change suggest that SOV might be a more basic order for gesturing events, but the present results show that while promising, the gesture task does not directly support this theory. The gesture task cannot be taken as an unbiased window onto the underlying orders that people use to conceptualize of and communicate events.

Where do these results leave noisy-channel theories of communication? The present studies attempted to test a specific model of rational production, assuming a simple noise channel that might disrupt the order of simple sentences. The data in these three experiments do not provide conclusive evidence either for or against this model, but do support the existence of an alternate explanation that cannot yet be ruled out. Noisy-channel models have been successful in predicting and explaining phenomena in natural languages at the level of phonological, lexical, and syntactic structure (Gibson, Bergen, et al., 2013; Jaeger, 2010; Levy et al., 2009; Mahowald, Fedorenko, Piantadosi, & Gibson, 2013). The central patterns of gesture production have been taken to mean that noisy-channel models might also explain basic word order phenomena (cf. GP2013), but the present research shows that this evidence is far from unambiguous.

It is possible that noisy-channel type effects do influence ad-hoc communication, but that gesture paradigms do not reveal them. The gesture task is attractive as a research tool because it is natural for participants and appears to be relatively free of native-language bias, as well as being rich enough to support a wide range of hypotheses and research questions. However, that same richness has a clear down side: participants inventing gestures and stringing them together have



the potential to reach new solutions to production pressures, and to combine theoretically separate aspects of the communication system in unexpected ways.

To solve these problems, we will need to move away from the task as it is currently administered. One possibility is to attempt communication games in more restricted channels of the types that have been used to study aspects of language evolution in the lab (cf. Kirby, Cornish, & Smith, 2008). However, it remains to be seen whether tasks like this will be rich enough for participants to design utterances to account for noise, without being transparent enough that they consciously strategize about the ordering or other features of their communication attempts.

Tasks like these, however, provide an important complement to studies of existing and newly developing languages as we attempt to understand the typology of word order and the role of communication effects in the design of language. Particularly for explanations of human language that draw on more general principles of robust and parsimonious communication, controlled lab studies offer the chance to test specific predictions of a theory with a high degree of control. Careful attention to the constraints and affordances of the particular mediums chosen when designing such ad-hoc communication experiments will make it possible to test these theories more precisely.

# CHAPTER 4 - ADULTS AND 5-6-YEAR-OLDS MAKE INFORMATION-SENSITIVE CHOICES ABOUT ARGUMENT DROPPING IN TRANSITIVE SENTENCES

## INTRODUCTION

Language use requires constant choices about what information to include or exclude. Even for very simple messages describing simple events, it is not always necessary or desirable to fully describe an event: if someone asks me *What are you doing?*, answering *I am eating* may be sufficient (compared to a longer alternative like *I am eating a sandwich* or *I am eating a grilled cheese sandwich*). Successful communication requires cooperation between a speaker and a listener, such that both will make the same assumptions about what a speaker means by the choices they make. Grice (1975) famously codified these conversational assumptions as a series of ‘maxims’, including the maxims of Quantity (‘give as much information as is needed, but no more’) and Relevance (‘say something that furthers the goal of the conversation’). Thus for instance *the grilled cheese sandwich* may indicate that the longer description is necessary, e.g. because there is another sandwich in the shared context.

While these kinds of pragmatic abilities have been most thoroughly studied at the phrase and sentence level, recent work has shown that adults implicitly design their speech at multiple levels to enhance communication. Humans are exquisitely sensitive to the statistical regularities of language in comprehension: they are sensitive to language-like structural regularities in infancy and use statistical information to predict word and structure choice during comprehension (Levy, 2008; MacDonald, 2013; MacDonald, Pearlmutter, & Seidenberg, 1994; Saffran, Aslin, & Newport, 1996). In a complementary way, language *production* is shaped to improve the chances of successful communication. Formalizations

based on information theory (Shannon, 1949) have been successful at quantifying how linguistic signals express information, explaining a wide variety of language phenomena including the distribution of word lengths in the lexicon and on the use of nonliteral language such as irony and exaggeration (Kao, Levy, & Goodman, 2013; Kao, Wu, Bergen, & Goodman, 2014; Piantadosi et al., 2011). Central to these models is a notion of a tradeoff between *cost* and *effectiveness*, recasting the maxim of Quantity—producers will attempt to say as little as possible while still conveying the message. The principle of uniform information density (UID) states that the optimal solution to this tradeoff can be reached by attempting to match the amount of (formal) information carried by each unit of language (Levy & Jaeger 2007 cf. T. F. Jaeger, 2013). In other words, people tend to spend more time producing the ‘hard parts’, providing extra linguistic information to ease the comprehension of less predictable elements. This takes place at multiple levels of representation, including phonological reduction, lexical choice (e.g. between *math* and *mathematics*) and inclusion of optional arguments (Aylett & Turk, 2004; Jaeger, 2010; Mahowald et al., 2013; Resnik, 1996; van Son & van Santen, 2005).

Notably, most of this work has focused on within-language informativity: the relative predictability of linguistic alternatives is calculated in terms of the global and conditional frequencies of sounds, words, and word combinations, or from estimations of conceptual content or predictability (e.g. that in general, a person is more likely to *eat an apple* than to *eat a door*, Maurits, Navarro, & Perfors, 2010; Resnik, 1996). A related body of work has focused on how referring expressions for objects (*that*, *that sandwich*, *that grilled cheese sandwich*) are related to a particular context, formed either by information from earlier in a discourse, or by the NON-linguistic information available in a scene (cf Sedivy, 1999). For non linguistic contexts, most studies have made the notion of informativity very concrete: a noun phrase like *my sandwich* or *my grilled-cheese sandwich* is informative, under-informative, or over-informative to the extent that the expression can be used to uniquely identify one of several referents in a context. The most important kind of context in this framework is the common ground, or information that is shared between both the comprehender and producer. Adults are skilled at correctly identifying objects to a communicative partner when there are possible confusable referents, but even they are not ‘strict’ in their use of common-ground information, looking to referents that their communication partner cannot see during early language comprehension of sentences like *the small candle* (Keysar, Lin, & Barr, 2003).

Children, on the other hand, appear to have trouble with appropriately informative referring expressions as late as eight years old. Children under the age of eight are often under-informative in their own speech, using definite noun phrases (*that one*) and pronouns (*she did it*) even when the listener cannot tell what is being referred to (cf. R. Brown, 1973). In some experimental settings, 5-year-olds fail to uniquely identify

objects, referring e.g. to “the triangle” when there are two triangles (Sonnenschein & Whitehurst, 1984). However, in other tasks, children do show evidence of understanding common-ground context; for instance, they describe objects (e.g. a big cup) in an array differently depending on whether or not their partner can see a potentially confusing referent (e.g. a small cup, Nadig & Sedivy, 2002). From toddlerhood, children are clearly sensitive to something about what other people know, and can use this information to solve other communicative tasks such as what a novel word refers to (Akhtar, Carpenter, & Tomasello, 1996). In addition, subtler online methods indicate that children as young as four are already be capable of *noticing* when a referring expression is ambiguous or unambiguous, as shown by eye movements and response speeds in action and comprehension tasks (Morisseau, Davies, & Matthews, 2013; Nadig & Sedivy, 2002; Nilsen, Graham, Smith, & Chambers, 2008). One of the central goals of this paper is to better understand the nature and extent of children’s pragmatic abilities during this period of paradoxical sensitive awareness and frequent failure in practice.

Frank & Goodman (2012) have developed a model of pragmatic inference that attempts to explain informative choices of referring expressions from the perspective of information theory. Critical to the model is a definition of a ‘rational speech act’, the idea that a speaker will choose their word/words to maximize the information that will be transferred to the listener. One approach to testing this model has involved reduced communication tasks, where people use a single word to identify a referent out a context set. For instance, in a context with a blue circle as a target with a blue square and a green square as a distractor, CIRCLE is preferred to identify the target shape. And in contrast, saying BLUE leads people to believe the blue square is being referred to (because if they had meant the blue circle, they would have said CIRCLE). Even three-year-olds have shown sensitivity to successfully understand referring expressions that require inferences about the intent of speakers; when possible referents form an ad-hoc scale (e.g. one bare smiley-face, one with glasses, one with glasses and a hat), children know that “My friend wears glasses” probably refers to the second one; if the speaker had meant to refer to the third face, she could have said “My friend has a hat” (Stiller, Goodman, & Frank, 2015).

This kind of task is very helpful for understanding whether and how participants calculate the informativity of potential utterances, because utterances limited to a fixed number of words can hold the *cost* of utterances constant (assuming that the words have, on average, the same costs of production) while varying either linguistic or nonlinguistic context. It is not currently known whether children can design utterances like “my friend with glasses” *for production*; a clearer understanding of what kinds of informativity calculations adults and children can use for understanding will put constraints on how we understand success (and failure) in production.

One such success or striking ability that has received relatively little attention in the pragmatics literature is the inclusion of optional arguments (I'm eating, I'm eating a sandwich). Verbs vary in their *selectional restrictions*, or what kinds of entities can appear as the subject, object, indirect object. These restrictions can be broad or narrow; I eat \_\_\_\_\_ will almost certainly be followed by a food (relatively high predictability/low information) while I see \_\_\_\_\_ places few restrictions on the identity of the object. Resnik (1996) modeled argument inclusion in terms of the conceptual content (taken from a general concept taxonomy) of arguments appearing with specific verbs, showing that the tendencies of verbs to include or omit arguments could be explained by the selectional restrictions on their arguments. Likewise, Maurits et al. (2010) used adult judgments of possible events made of words from child directed speech to create a database of possible event likelihoods. They showed that word order taxonomies may be at least partially influenced by uniform information density: each word contributes new information to an unfolding sentence, and some word orders turn out to provide, on average, smoother information profiles than others (e.g. by 'spreading out' unlikely elements).

How might argument inclusion be affected by specific context, in addition to broader language statistics or conceptual knowledge? As with simple referring expressions, participants must determine what (if anything) is necessary to say to pick out a referent from a set. But in addition, they must potentially keep multiple sets in mind, especially when making a trade-off with the cost of an utterance. That is, there will be a separate (though possibly overlapping) set of possible referents corresponding to each argument position of a verb. The problem faced by children (or adults) deciding how to produce or comprehend a sentence with argument drop is therefore more complex: the size of the context set of *events* is the *product* of possible subjects, objects, and any other arguments. Do adults and children use non-linguistic context in this way?

There is evidence that young children are aware of *argument structure*, the basic who-did-what-to-who information carried in the structure of a sentence, from early in development (cf. Fisher, Gertner, Scott, & Yuan, 2010). However, even if children have a good understanding of argument structure when expressing a single event, they face additional challenges if considering the possible set of alternate events. Many verbs allow unspecified objects, and while subjectless sentences are generally ungrammatical in English, they are fully grammatical in pro-drop languages like Spanish, where a sentence like *Como empanadas* is acceptable as a standalone description, not just an elliptical response to a question. Deciding which arguments can be omitted would require determining not just what other referents are in the context, but transforming that into a representation of the other possible events, taking into account which of them might have been the agent of an eating event (e.g. *myself, my brother*) and

which might have been a patient (e.g. *a banana, a sandwich*.) Even for adults, we do not currently know whether inclusion or omission of sentence arguments tracks with the alternate events their listener might be considering.

The present studies ask how adults and children use this information when their choices about argument inclusion are restricted. In three studies, we provide and test a novel paradigm for studying adults' and children's abilities. This paradigm has two key parts. First, in both comprehension/evaluation and production tasks, we require participants to work with transitive sentences (e.g. *The girl reads the book*) that are described in just two words (e.g. GIRL-READ or READ-BOOK). We expect that participants will usually choose to include verbs, because their informativity will be kept both constant and relatively high (because listeners have uncertainty about *what* the set of possible event participants will do.) In addition, we deal with cases where participants are considering a set of people (possible agents) and a set of objects from a small semantic class (e.g. fruits, farm animals). Thus, additional possible agents or patients do not significantly increase the information about possible selectional restrictions of the verb: EAT is just as likely if the object is {apple} as if it is in {apple, banana}. For this reason, we focus on the choice to include *either* the agent or the patient of the event, but not both. Thus the cost of the utterance is held constant (assuming that the lexical items are all about as easy to retrieve/pronounce) while allowing argument choice (e.g. subject or object) to vary: the decisions are not about *how much* to say but *what* to say given limited resources.

The second key part of the paradigm is *contrast sets* that contain groups of potential agents and/or patients of an event (See Figures 1 and 3 for examples). Here, we assume that both adults and children will be able to identify referents that are more likely to be subjects or objects; we are not testing whether participants know that a girl is more likely to be the subject of READ than a book is. Rather, we test whether they implicitly calculate the *possible events* that correspond to a contrast set when deciding which of the possible two-word phrases conveys the most information. By varying the sets of possible agents and/or patients and limiting descriptions (in production or comprehension/evaluation) to only two words, we create a paradigm in which the same phrase can express varying amounts of novel information about an unknown event depending on the context. For instance, in the leftmost array in Figure 1, the full message to be communicated might be *The boy reads the book*. The array corresponds to two possible events, the boy reading the book or the boy reading the newspaper. Assuming that the verb is included, there are two possible messages (BOY READ or READ BOOK). For this array, BOY READ fails to resolve the ambiguity: a listener can probably determine that the boy is the agent, not the patient of the

event. On the other hand, READ BOOK eliminates the ambiguity by specifying the patient and relying on an intelligent listener to identify the unique possible agent from the array. However, if the context array contained a girl instead of the newspaper, the reverse would be true: when the possible events are the boy or the girl reading the book, BOY READ becomes the sentence that can eliminate the ambiguity.

The goals of this paper are to show that this kind of inference is possible for both adults (indicating that it may help to explain argument choice in adult speech) and children (indicating that such inferences may play a role in early language use). This will clarify our fundamental pragmatic abilities, extending informativity-based models of language to incorporate non-linguistic context and explain basic argument structure choices. In Experiments 1 and 2, adult participants are given space for just two words to try and communicate a secret event to a communication partner who can see an array of potential agents and patients. In Experiment 3, five- and six-year-olds hear two puppets attempt to describe hidden events for the child's parents. We provide two speakers rather than having children give ratings of a single sentence both in order to make the task comprehensible for younger children who may struggle with rating scales, and to make sure children are comparing the relevant alternatives (e.g. MONKEY EAT and EAT ORANGE, rather than the more complete MONKEY EAT ORANGE). These studies provide a novel tool for understanding how children evaluate the informativity of sentence structures and establish a baseline for further study of children's developing pragmatic skills.

### EXPERIMENT 1 – ADULT PRODUCTION WITH '2 VS. 1' CONTRAST SETS

When faced with grammatically appropriate alternatives and referential communication tasks, adult speakers select words and constructions to make utterances easy to understand (Piantadosi, Tily & Gibson 2011; Levy & Jaeger, 2007, Frank & Goodman 2012.) Do they do the same with telegraphic speech when choosing whether to express the agents and patients of an event? To begin answering this question, we establish whether adults can make appropriate inferences like this in an internet-based task. We use the simplest cases (i.e. two alternatives for one thematic role, and one for the other), presenting the task to participants as a study designed for children in which they tried to explain the events to an onscreen cartoon character. To our knowledge, this type of 'human simulation' paradigm of limited production has not previously been attempted, though it is inspired in part by simulations of comprehension (Gilette et al.,

1999)<sup>7</sup>. We predicted that when there were fewer agents than patients, participants would be less likely to mention subjects: in Figure 1, mentioning the agent in a response (e.g. BOY READ) leaves two possible events, while mentioning the patient (READ BOOK) reduces the uncertainty to just one event.

## METHODS

**Participants** 38 English-speaking adults participated on Amazon’s Mechanical Turk. Participants were screened to be located in the United States and self-reporting English as their first language (an additional 6 participants were excluded who did not meet these criteria.) No other demographic information about participants was collected. The task took approximately 6 minutes and participants were paid \$0.75 for the study.

**Stimuli** We created sets of cartoon stimuli for each of six simple, early-acquired verbs (*eat, feed, drink, wash, read, and roll*). Each set consisted of an exposure and action picture for the Object-Ambiguous and Subject-Ambiguous condition. Exposure pictures showed the agent and patient of the action at rest, along with either an additional possible agent or patient. They also included an image of a cartoon character sitting in a chair off to the side. The ordering of agent, patient, and distractors was varied between verbs for variety, but kept constant within each verb to control for biases to attend to center stimuli. Action pictures showed the same chair, empty, along with the agent and patient performing the action, and the distractor visible in the same position as in the exposure picture. An example stimulus set is shown in Figure 1, and all stimuli are available at [http://tedlab.mit.edu/~mekline/Stimuli/SubDrop\\_SDPictures/](http://tedlab.mit.edu/~mekline/Stimuli/SubDrop_SDPictures/).

All stimuli were presented to participants online using Python and the EconWillow package (Weel, 2008). Code for running the experiment can be found at <https://github.com/mekline/Subject-Drop>.

---

<sup>7</sup> Concurrently with this experiment, our lab has also tested whether adults include or omit lexical items that are very predictable or very unpredictable from the verb (e.g. the *policeman arrested the plumber*, where *policeman arrested* is highly predictable but *arrested the plumber* is relatively surprising.) This study found results supporting the hypothesis of this paper; adults were more likely to include arguments that were *unpredictable* from the verb (Mahowald, Kline, Fedorenko & Gibson 2015).



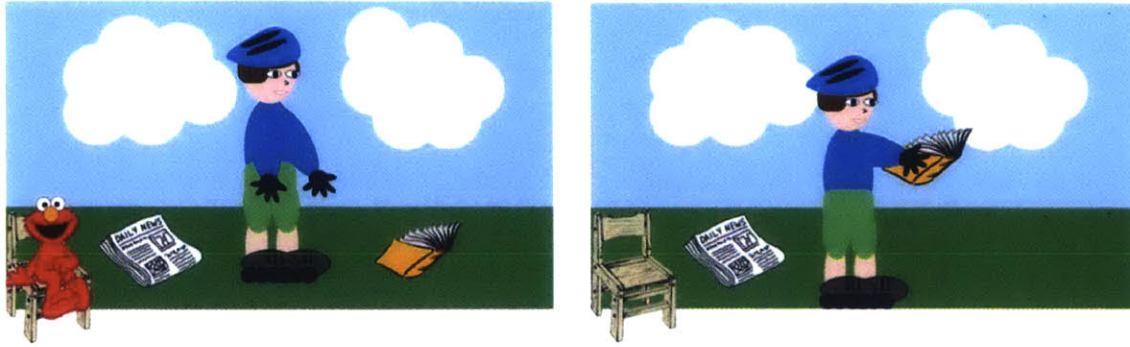


Figure 1: Exposure and action stimulus photos for the “reading” event in the Object-Ambiguous condition. Elmo is present for the presentation of the possible referents in the action, but absent when the action is shown. The Subject-Ambiguous versions were identical except that a girl was present instead of the newspaper.

**Procedure** After confirming their consent and ability to see pictures in the online testing environment, participants were told that they would be talking with Elmo about a series of cartoon pictures, but that he would run away each time and miss the action that took place. They were told that they could only send two words to Elmo at a time, but still had to try and convey what was happening in each scene. Participants saw six trials (one per verb) in a random order, including three each in the Object-Ambiguous and Subject-Ambiguous conditions. On each trial, participants first read a sentence corresponding to the appropriate condition describing the items they would see in the scene (e.g. “In this scene, there is a grandma, a baby, and a ball”), and then saw the exposure picture of these items (with Elmo in his chair viewing the display) for 5 seconds (timing was controlled so that participants could not skip through presentation of the videos). They then saw the action picture for 5 seconds, showing the agent performing the action with the patient, with the distractor item visible, and Elmo’s chair empty. The image was paired with a descriptive transitive sentence e.g. (“The baby rolls the ball”.) Finally, the exposure picture reappeared, with instructions to tell Elmo what had happened. Participants were given two text boxes to enter their response in.

If participants entered something other than one word in each text box (e.g. writing “baby rolls” in the first box and “ball” in the second), they were told to try again. The number of words was determined simply by checking that at least 2 characters had been entered and that no spaces were used. No other restrictions were placed on the answers that participants could enter. After participants gave a 2-word answer, they saw a progress bar, and were told both the time it had taken them to enter a response on that

trial, and their average speed so far. This was done to encourage participants to answer as quickly as possible.

After seeing all six trials, participants were told about the purpose of the study and given a unique code that they entered on Mechanical Turk to receive payment.

## RESULTS

Responses were first checked for minor variations in response form such as capitalization and verb form (e.g. “Eaten” was recoded as “eat”, but “hungry”, a word that did not appear in any of the given descriptions, was retained.) Some participants gave responses that were intended to convey a more general idea of the scene (e.g. “Monkey hungry”), but the majority of responses (80%) consisted of two of the possible three content words in the sentence (subject, verb, and object.) Because of this variation, we

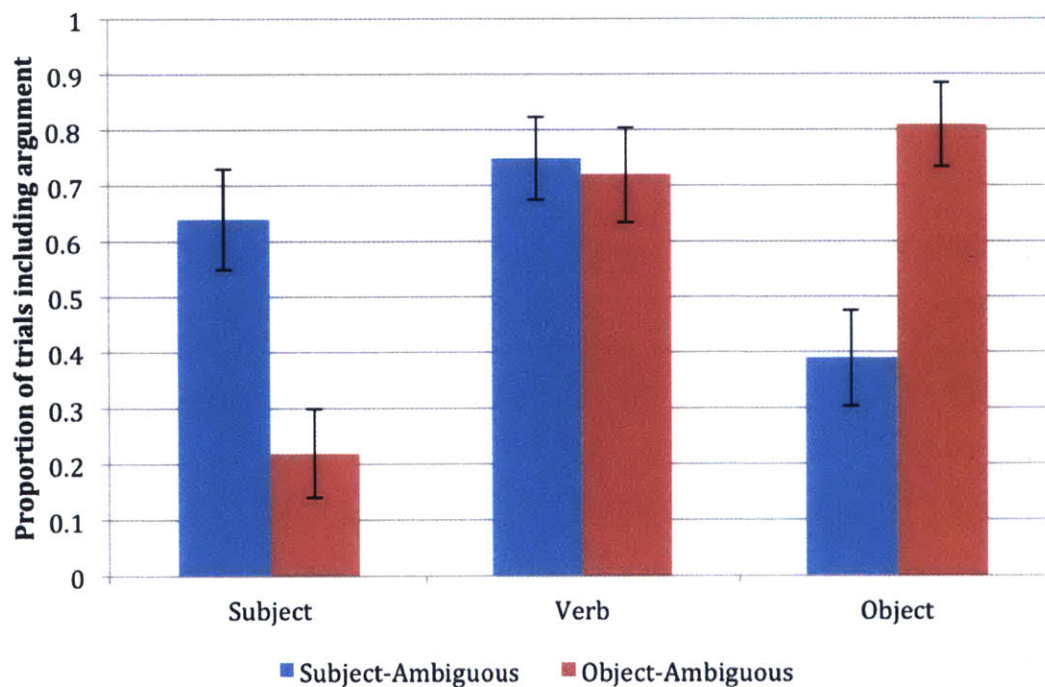


Figure 2: Results of Experiment 1, showing the proportion of trials on which participants included the subject, verb, and object of the sentence in their response. Error bars represent bootstrapped confidence intervals around the mean.

analyzed our data for responses that contained the subject or object, rather than responses of the exact form SV/VO<sup>8</sup>.

The results of the experiment are summarized in Figure 2. When participants saw arrays with an ambiguous object, they were more likely to mention the object in their response, and when they saw an array with an ambiguous subject they were more likely to mention the subject. The effect of condition (Object-Ambiguous or Subject-Ambiguous) on whether participants mentioned the subject in their response was highly significant by a mixed-effects logistic regression with random slopes and intercepts for item and random intercepts for participant ( $X^2 = 9.16$ ,  $df = 6$ ,  $p < 0.01$ ). The same was true for mention of the object ( $X^2 = 11.77$ ,  $df = 6$ ,  $p < 0.001$ )<sup>9</sup>. We also found that participants overall were somewhat more likely to mention objects than to mention subjects (59% vs. 43%). A small percentage (11%) of answers consisted of both subject and object; excluding these, there were significantly more objects than subjects mentioned ( $p < 0.01$  by a binomial test.)

## DISCUSSION

As predicted, adult participants tended to include the most informative words in their constrained phrases, mentioning the subject of the event when the subject was ambiguous from the context array but omitting it when a listener could determine the likely subject simply by examining the array. Although not instructed to do so, participants usually included the verb in their response; this is likely due to the fact that the verbs were not explicitly represented in the arrays at all, making mention of the verb highly informative for a listener (verb mention did not vary across the two array conditions.)

The fact that participants were overall more likely to mention objects than subject (i.e. to produce V-O phrases) is especially interesting since it mirrors patterns of argument drop in both adult languages and in

---

<sup>8</sup> Restricting the data to only responses that consist of either exactly SV or exactly VO, we find no significant differences in the reported results.

<sup>9</sup> In both cases, the model with full random slopes and intercepts for both items and participants failed to converge. The full model specification for the results given above can be executed in R with the following code: `lmer(mentionSubject ~ trialVersion + (1|Paycode) + (trialVersion|verb), data=mydata, family="binomial")`.

language acquisition. This study cannot determine why adults were more likely to produce responses like *reads book* than *boy reads*; it might be due to a preference to produce complete (VP) constituents, a general bias to include words later in the sentence, or simply to a difference in salience in our materials.

Another limitation to this first study is that the calculations necessary are very simple: adults might determine what to say simply by noting that one argument is ambiguous and one is not. That is, rather than by calculating the space of possible events and determining which arguments reduce uncertainty by the greatest amount, this particular task could be solved if adults are able to notice whether a possible phrase uniquely identifies the target event, but not to calculate and compare informativity more generally. Experiment 2 therefore examines this more extended ability.

## EXPERIMENT 2 – ADULT PRODUCTION WITH ‘MANY VS. MANY’ CONTRAST SETS

The second experiment was designed to test whether adults in fact calculate the relative informativity of possible utterances, even when those utterances do not uniquely identify a target event. This experiment is important for two reasons: first, establishing a clear computational-level description of adults’ abilities will provide important context for testing the development of these abilities. Second, this manipulation can strengthen evidence for rational-actor models of language by providing another test case in which the most pragmatically appropriate thing to say is not determined simply by whether or not a referent is uniquely identified out of a context, but rather by a more robust notion of informativity which takes into account what else one might have said and which of a series of possible utterances can reduce uncertainty to the greatest degree.

To implement this test, we provided adult participants with larger arrays that presented sets of 7 potential event participants (see Figure 3.) In some cases, the event could be uniquely identified with a two-word utterance, but in other cases this was not possible, because there were both multiple agents and multiple patients present. However, we predicted that adults would still be able to make judgments about the most informative arguments to mention: when there were more agents than patients, we predicted that participants would be more likely to mention subjects. For the top right array of Figure 3, mentioning the agent in a response (e.g. JOHN FEEDS) leaves just two possible events (he feeds the dog or the duck) while mentioning the patient (FEEDS DOG) only reduces the uncertainty to five events (Amy/John/Emily/etc. feeds the dog.)

## METHODS

**Participants** 91 English-speaking adults participated on Amazon’s Mechanical Turk. Participants were screened to be located in the United States and self-reporting English as their first language (an additional 21 participants were excluded who did not meet these criteria.) No other demographic information about participants was collected. The task took approximately 13 minutes to complete and participants were paid \$1.00 for the study.

**Stimuli** For experiment 2, we created a series of images very similar to Experiment 1. We created sets of cartoon stimuli for each of twelve simple, early-acquired verbs (*eat, feed, hold, drink, kick, drop, wash, pour, throw, touch, read, and roll*). Each set consisted of an Action picture showing a person performing an action with an object, and a set of six possible Exposure pictures. Each exposure picture showed a group of possible agents (people) and possible patients (objects). The people were generated using a character-creation website (CITE Simpsons, no longer available ☹), with distinct features and names added to their shirts. The objects always consistent of members of a category that could be the patient of the given verb (e.g. various foods that could be eaten.) The number of possible agents and patients was varied between versions, and always included the actual agent and patient: there could be 1 agent and 6 patients, 2 agents and 5 patients, all the way to 6 agents and 1 patient. We use the abbreviation 6\_1 to refer to a scene with six possible agents and one possible agent. A full set of Exposure pictures and the corresponding Action picture are shown in Figure 3, and all stimuli are available at [http://tedlab.mit.edu/~mekline/Stimuli/SubDrop\\_MDPictures/](http://tedlab.mit.edu/~mekline/Stimuli/SubDrop_MDPictures/). All stimuli were presented to participants online using Python and the EconWillow package (CITE). Code for running the experiment can be found at <http://www.github.com/mekline/Subject-Drop>.

**Procedure** After giving consent, participants were introduced to the task. They were told that they were providing descriptions of simple scenes for another Mechanical Turk participant, but would only be allowed to give two-word responses. Each participant saw 12 trials in a random order, two at each exposure level (e.g. 6\_1, 5\_2, etc). On each trial, participants saw the exposure display for ten seconds, and then read the transitive sentence corresponding to the action they would see (e.g. “John feeds the dog”). They then saw the Action picture corresponding to this sentence (see Figure 3) for 10 seconds. Finally, the Exposure picture reappeared and participants were asked to provide a two-word description of the event in two text boxes. Response entry, control, and feedback were identical to Experiment 1.

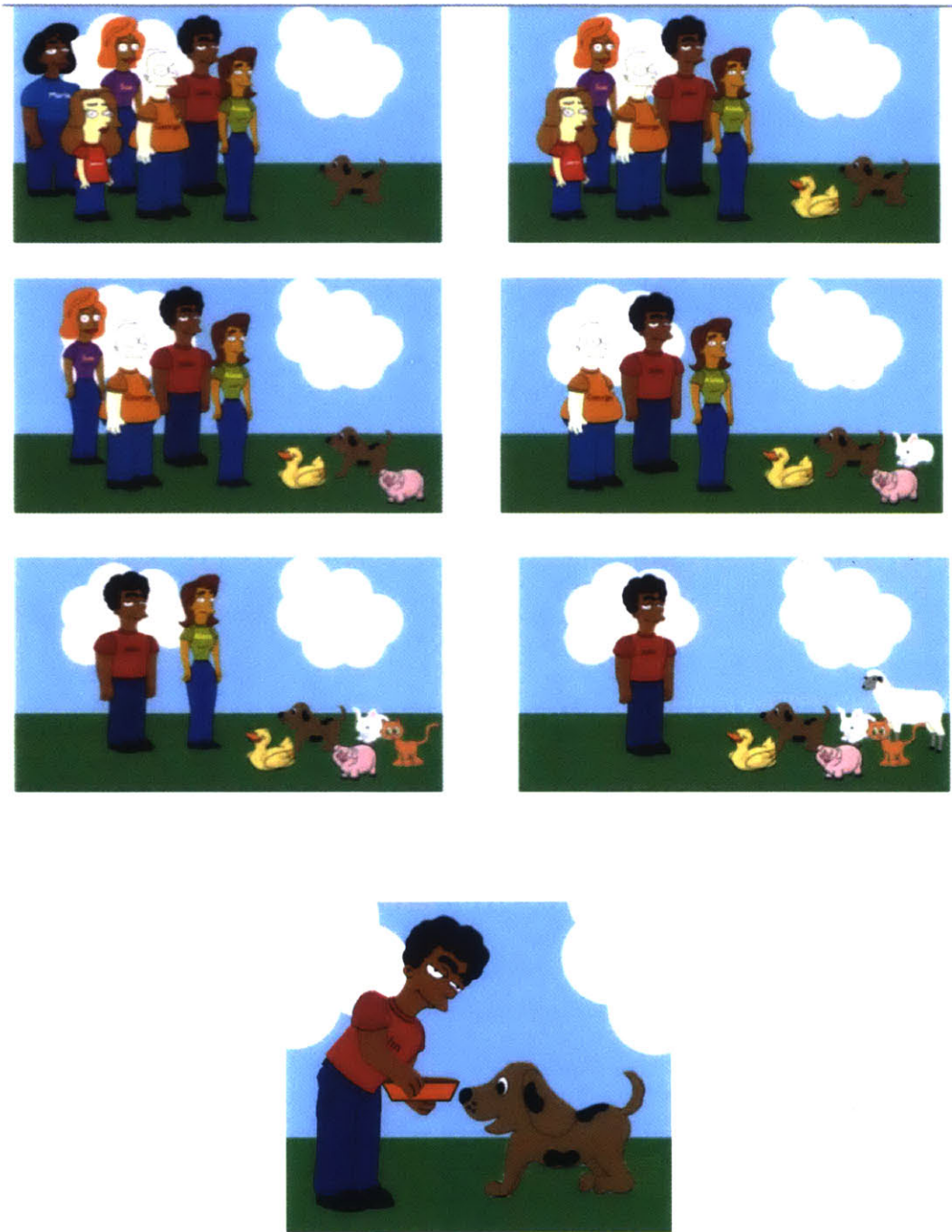


Figure 3: Exposure images used for the sentence “John feeds the dog” in Experiment 2. Each participant saw one version, showing a set of possible agents and patients.

## RESULTS

As in Experiment 1, participants' responses were first coded to identify uses of the three target words (Subject, Verb, Object) from each sentence. Again, participants sometimes introduced other words in their responses. In contrast to Experiment 1, all of the potential agents that participants saw were humans, and all of the potential patients belonged to the same superordinate category, and so participants sometimes used general descriptors or pronouns instead of the target nouns (e.g. *Man*, *she*, or *fruit*.) Responses were coded as referring to the subject or object *only* if the word used uniquely identified the referent out of the set of 6 possibilities, even if the participant happened to have seen only one category member (e.g., "woman" was not coded as referring to the subject of the sentence "*Mary ate the orange*" even if Mary was the only woman in the set of possible agents on that trial.) As in Experiment 1, participants sometimes used words outside the set of subject, verb, and object (14% of responses included an unexpected word). We therefore continued to use the presence of subject and/or objects in the

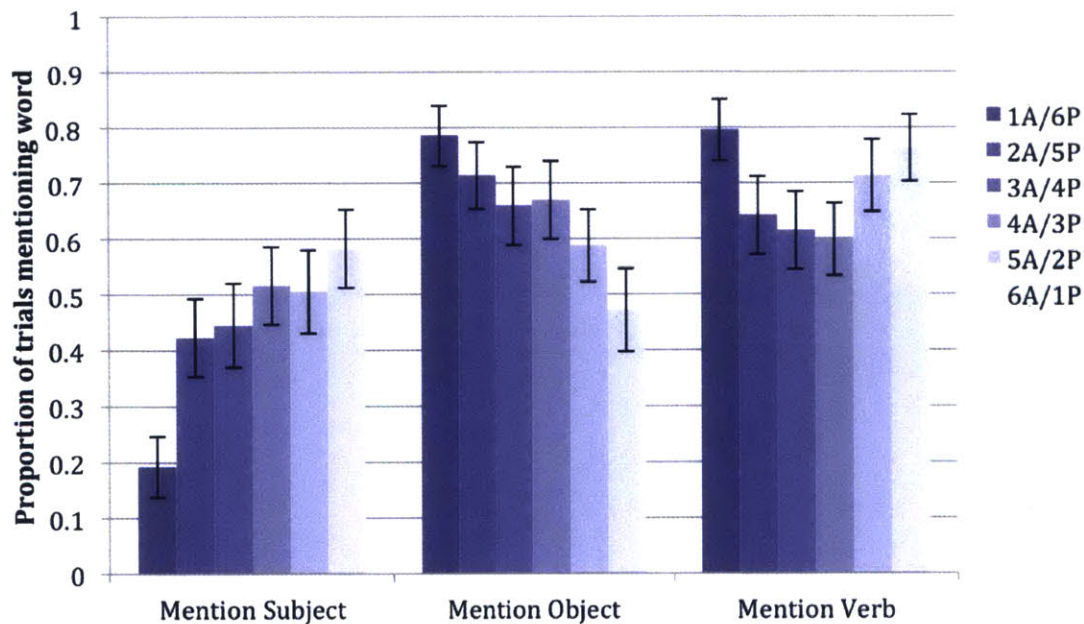


Figure 4: Trials on which participants included subjects, objects, and verbs. Error bars represent 95% bootstrapped confidence intervals.

response, rather than exact SV or VO responses, to analyze response patterns<sup>10</sup>.

The results of the experiment are summarized in Figure 4. The effect of array type on whether participants mentioned the subject in their response was highly significant by a mixed-effect logistic regression with random slopes and intercepts for both item and participant ( $X^2 = 12.3$ ,  $df = 8$ ,  $p < 0.001$ ). The same was true for mention of the object ( $X^2 = 17.7$ ,  $df = 8$ ,  $p < 0.001$ ). The patterns of these effects are as predicted (Figure 2) – as more agent distractors (and thus fewer patient distractors) were present, participants became more likely to mention the subject and less likely to mention the object (44% vs. 64%). On the subset of trials (73%) on which participants mentioned only a subject or object, there were significantly more objects than subjects mentioned ( $p < 0.001$ , binomial test).

To test whether participants in fact gave graded responses to the intermediate arrays (e.g. two agents, five patients vs. four agents, three patients), we also examined the effects of array type after removing from the dataset the arrays for which a ‘deterministic’ answer could be given (e.g. one agent, six patients or six agents, one patient.). Examining these four middle levels, the effects of array type on both subject and object mention were still significant (Subject mention:  $X^2 = 6.17$ ,  $df = 8$ ,  $p < 0.05$ ; Object mention:  $X^2 = 4.6$ ,  $df = 8$ ,  $p < 0.05$ ).

## *DISCUSSION*

As predicted, when participants described events after seeing arrays with a set of possible agents and patients, their two-word answers reflected the degree to which a given word could convey new information (i.e. by how much it reduced the set of possible events.) This ability was not limited to cases where an event could be uniquely identified by a two-word utterance: even for the ‘middle’ cases where there were multiples of both agents and patients in the array, participants still traded off mention of subjects and objects to choose the arguments that reduced uncertainty to the greatest degree possible.

In addition, we replicated the effect of object-mention found in Experiment 1: adults were more likely to produce verb-object sequences overall. This replication raises the possibility that this effect is related to typological and acquisition patterns in language more broadly, as opposed to being an artifact of a particular stimulus set.

---

<sup>10</sup> All results reported are qualitatively similar if only SV and VO responses are included in the analyses.



Having established that adults are able to make this novel kind of pragmatic inference in this telegraphic language task, we next ask whether young children are able to make the same kinds of inferences.

### EXPERIMENT 3 – CHILD COMPREHENSION/SPEAKER CHOICE WITH ‘2 VS. 1’ CONTRAST SETS

While performance varies across tasks, 5- and 6-year-olds are generally agreed to be sensitive to some aspects of pragmatic inference, though they may still be struggling to deploy this reasoning in particular cases such as scalar implicatures (Davies & Katsos, 2010; Huang & Snedeker, 2009; Nilsen et al., 2008; Stiller et al., 2015). We therefore begin with this age group, asking whether these children understand the relative informativity of utterances not just for the case of identifying a noun referent, but for the case of events as well.

The goal of the present study was to make it as straightforward as possible for children to understand the task as presented. While the materials were adapted somewhat for children (into a storybook which is described below), the structure of the experiment was parallel to Experiment 1, with either two agents and one patient, or two patients and one agent. We used two competing speakers on each trial who both attempted to describe a transitive event with just two words, rather than asking children to produce such sentences themselves. By providing two speakers/sentences to choose between, rather than one sentence to be rated, we hoped that children would be better able to appreciate the relative contrast, and that they would find the task natural. Conscious of the possibility that children are generally tolerant to pragmatic violations (Davies & Katsos, 2010), we described the speakers in each case as either “really good” or “just okay” at describing the scenes. A final decision for this initial study was whether the ‘listener’ who did not know about the events should be the child themselves or a third party; both methods have been used to evaluate what children know about whether sentences are informative. In the present version, children were allowed to see the hidden event, and heard the puppets describing the scene to their parents; this required them to keep in mind what their parents did not know. Future work will be able to determine what role this kind of theory-of-mind judgment might play in the development of pragmatic abilities.

### *METHODS*

**Participants** 23 monolingual (no more than 25% input in another language) children between the ages of 5;0 and 6;11 participated at the Boston Children’s Museum. Data from 4 additional children is not included because the child did not complete the experimental session.

**Materials** A set of stimuli very similar to those used in Experiment 1 were adapted into storybooks for use with children. The storybooks consisted of two ‘surprise event’ trials presented in a fixed order (*Eating* and *Feeding*). Each trial booklet involved a target event (e.g. the girl feeds the dog) and possible distractors for both the agent and the patient (e.g. the boy/girl feeds the dog/cat). Each trial started with pictures of the four possible event participants on successive pages, followed by images of the four possible events; the position of the target event in this series was varied across verbs.

The critical manipulation was created with two versions of an array showing the agent of the target event, the patient, and a distractor. Participants were assigned to either the Subject-Ambiguous or Object-Ambiguous condition, corresponding to two versions of the storybook that differed only in the content of these arrays on each trial. An example of the Subject-Ambiguous and Object-Ambiguous arrays along with the target event is shown in Figure 5. The arrangement of the agent, patient, and distractor was varied between verbs, but fixed for both versions of the array. After this was the final page of the trial, which showed the array again, next to a flap of paper covering the picture of the target event.

In addition to the storybooks, the experiment involved sets of matched ‘muppet’ cutouts presented on popsicle sticks mounted on a base. For each trial, two similar characters dressed in distinctive ways (e.g. striped shirt and polka-dot shirt) were created to serve as the two possible communicators.

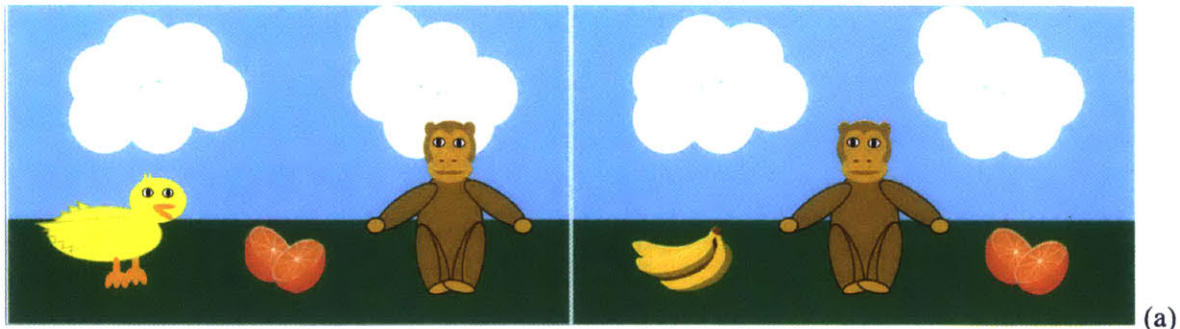
**Procedure** Participants were told that they would be reading a set of story books, with help from a series of puppets:

*Here’s my two muppets who are gonna help us tell some stories. These muppets are from another place, so they don’t speak English that well. They can only say one or two words at a time, so we have to listen to them very closely. One of them says “Me muppet!” and the other one says, “Me muppet too!”*

*So there’s two of them! One of them’s gonna do a really good job telling the story, and one’s just gonna do an okay job, all right? I want you to help me decide which one does a better job.*

The muppets were described as ‘really good’ and ‘just okay’ to help make the eventual judgment more natural for the children (since the speakers in the study only ever made true statements about the events).

The experimenter then proceeded to lead the child through the first storybook. The experimenter labeled each potential participant in the event (e.g. monkey, duck, banana, and orange), and then labeled the pictures of the possible events (*Look, here the duck is eating the orange.*) The experimenter attempted to use neutral prosody for each event, and was careful to use full nouns when they repeated between scenes (e.g. *Now, the duck is eating the banana*, rather than *Now, he’s eating the banana*). Sometimes



(b)



(c)

Figure 5: Subject-Ambiguous (a) and Object-Ambiguous (b) arrays presented to children in Experiment 3. These arrays correspond to the target event & sentence “The monkey eats the orange” (c).

participants started to label the images they saw; the experimenter said “Yeah!” or corrected the child if they mis-identified the scene.

After the four possible events had been presented, the experimenter showed the array (either Subject-Ambiguous or Object-Ambiguous) to the child and pointed out the ambiguity:

*Ohh, look, now what do we have here? Now we have three things! Look, there’s (a duck)/(some bananas), some oranges, and a monkey. Do you want to find out what happens next?*

Note that in this script, the experimenter does not describe the specific ambiguity (e.g. *there’s two fruits*) or describe the possible events corresponding to the list of entities. The experimenter then proceeded to the target page, which showed the array again, and a flap of paper covering the target event. The back of the storybook was lifted and oriented so that the experimenter, the child, and the two muppets could see

this page, but the child's parent/guardian could not (*Oh, see this flap here? Your mom doesn't know what's under there. You'll get to see what's under there in a sec, but your mom doesn't know!*)

The child was invited to lift the flap and see the picture of the actual event (e.g. *The monkey eats the orange*). The experimenter confirmed with the child that they knew what happened, and brought each muppet over separately to look at the picture as well. They then instructed the child to close the flap, and laid the storybook back on the table.

The experimenter reminded the child that both puppets would 'try and tell the story', and then spoke for each of the two puppets:

*This muppet over here in the polkadots says MONKEY EAT. And this muppet over here in the stripes says EAT ORANGE. Okay? So this one says MONKEY EAT, and this one says EAT ORANGE. So, which muppet did a better job telling your mom the story?*

The child pointed to one of the two puppets (if they did not do so spontaneously, the experimenter asked them, *Can you point to one of the muppets?*). Participants sometimes spontaneously described the event instead of choosing a puppet, either repeating one of the muppets (*Monkey eat!*) or describing the whole event (*The monkey ate the orange!*) In either case, the experimenter responded by asking the child to touch the one who did a better job telling the story. In all cases, the experimenter did not provide explicit positive or negative feedback, but said "okay!" in a positive tone and moved on. After this, the experimenter introduced a new set of muppets (so that opinions about which muppet was most informative would not be carried between trials). On both trials, the muppet who mentioned the subject spoke first, but the position of the muppet speaking first (left or right) was balanced across the two trials.

Children's points were coded offline by a research assistant with the child's first clear point directed at a puppet taken as their choice. 30% of experimental sessions were re-coded by the experimenter; agreement was at 95%.

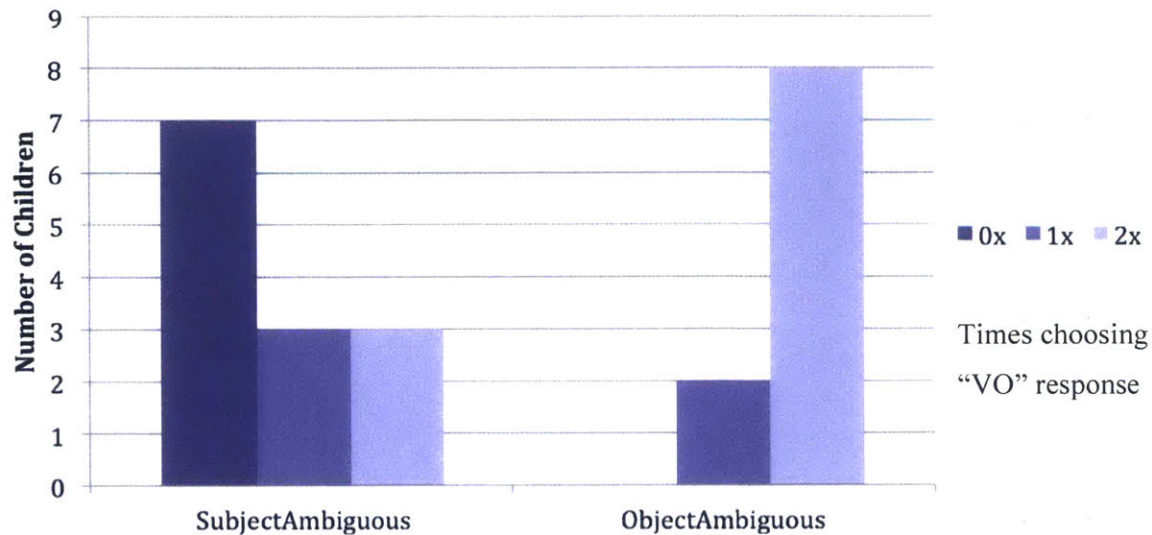


Figure 6: Number of times children chose puppets mentioning the object (EAT ORANGE) or subject (MONKEY EAT) in Experiment 3

## RESULTS

The results of this experiment are shown in Figure 6. When children saw arrays with an ambiguous subject, they were more likely to choose the muppet who had mentioned the subject (MONKEY EAT) as the informative speaker, and when they saw arrays with ambiguous objects, they were more likely to choose muppets who mentioned the object (EAT ORANGE.) The effect of condition (Object-Ambiguous or Subject-Ambiguous) on whether children chose the puppet mentioning the object was significant by a mixed-effects logistic regression<sup>11</sup> ( $X^2=12.22$ ,  $df=4$ ,  $p<0.001$ ).

---

<sup>11</sup> The full model specification for this regression is given in R code by `lmer(choseObjectDrop ~ Condition + (1|trial) + (1|Subject, data=sub.long, family="binomial"))`. Random slopes for trial were attempted but the model did not converge.

## DISCUSSION

In parallel to adults, 5- and 6-year olds judged two-word descriptions of transitive events according to their relative informativity. When there were two possible agents in an array and just one possible patient, children preferred speakers who mentioned the subject of the sentence, disambiguating the target event. In contrast, when the arrays contained a single agent and two possible patients, children preferred the speaker who clarified which was the undergoer of the event.

As discussed in Experiment 1, there are two possible interpretations of this experiment. First, children might succeed in this task by calculating the relative informativity of each phrase and determining that one phrase successfully identifies the target event and one does not. This would still require them to perform the more general calculation about informativity, representing the possible set of events that could have occurred and evaluating the amount of information that is transferred to the listener by each phrase. However, we do not yet know whether they are able to compare the informativity of statements which are both partially ambiguous (i.e. partially but not fully informative.) Additional testing with larger arrays (parallel to Experiment 2) may be able to distinguish between these hypotheses.

## GENERAL DISCUSSION

Across three tasks in a novel paradigm, we show that both adults and 5- 6-year-olds consider the informativity of sentences with dropped arguments when they engage in communication tasks. Although the two kinds of shortened sentences (Subject-Verb, e.g. *Boy-Reads*, and Verb-Object, e.g. *Reads-Book*) are equal in length (cost) and express the same amount of information about the target event in a neutral context, adults and children both recognize that the informativity of these phrases depends on the set of possible alternative events. When there are more possible *agent* alternatives in context (i.e. when the subject is ambiguous), both participant groups prefer phrases that include the subject, and when there are more patient alternatives they prefer phrases that preserve the object.

These findings build on what we know about the pragmatics of referring expressions across development: in addition to tracking what possible alternative objects are visible in a scene, this task requires people to represent these objects in terms of their possible thematic roles in events, and to calculate the set of possible events that could be constructed from these roles. Our findings suggest that adults and 5- 6-year-olds use the relative informativity of constrained-length descriptions of transitive events to decide what to say (adults) or determine which of two speakers is the most helpful (children).

There are three directions for additional work with children. First as discussed above, testing children's beliefs about speaker informativity in the context of larger, more ambiguous arrays (in line with Experiment 2) will allow us to better understand the ways in which children evaluate ambiguity and determine the informativity of possible descriptions. Second, other pilot data we have collected indicates that this task can be straightforwardly extended to 3- and 4-year-olds, allowing us to chart the development of these abilities. In the case that these young children do *not* show a preference for more informative speakers, we will be able to run control conditions that contrast speakers who say true and false things (e.g. JOHN READ vs. MARY READ for a target event of John reading the book) to make sure that children are willing and able to identify more competent speakers in this task when the calculation is straightforward and requires no pragmatic inferences.

Third, this paradigm will make it possible to clarify the role of the development of theory-of-mind in children's pragmatic inferences. Young children are aware (especially when measured with sensitive online measures) when they themselves are given an uninformative answer to a problem – being able to detect this is a critical prerequisite for pragmatic inferences more generally. Broadly speaking, the earliest successes in pragmatic inference are found when the child is in the role of the comprehender and does not need to represent how much their conversational partner knows or doesn't know about a scene (Morisseau et al., 2013; Nadig & Sedivy, 2002; Stiller et al., 2015). Thus it is possible that the failures on other tasks, especially in production, may have to do with simultaneously tracking the *informativity* of sentences and the *informedness* of their conversational partners. In Stiller and colleagues (2015), children had to consider the possible descriptions the speaker might have used, and the models that have been proposed by the authors to explain success on this task involve nested representations of what speakers and listeners do and do not know<sup>12</sup>. However, children in this study did not have to explicitly consider a conversation partner who knows less than they themselves do, as is the case both in Experiment 3, and in everyday language production. This distinction is potentially important for the development of pragmatic abilities,

---

<sup>12</sup> C.f. Frank & Goodman (2014), listeners are described as modeling a speaker's knowledge state, what the speaker knows about what the listener knows: "by assuming that speakers choose their words to be informative in context, listeners routinely make pragmatic inferences that go beyond the linguistic data".

since there is an established dissociation between implicit and explicit use of theory-of-mind information by young children (Onishi & Baillargeon, 2005; Wimmer & Perner, 1983).

In a sense, children in Experiment 3 are solving a harder task because they must remember that their own parent knows some, but not everything, about the events at the same time as they recall what they (and the puppet speakers) are learning about the target event. We have also piloted a version of this study which is more similar to previous studies of comprehension of over- and under-informative utterances, in which it is the *child* who doesn't know what the target event is (and they are then asked to determine which puppet helped them figure out the target event.) In non-linguistic tasks, children are skilled at detecting competent and helpful people even in infancy (Hamlin & Wynn, 2011; Kuhlmeier, Wynn, & Bloom, 2003). Is 'helpfulness' related to the calculation of linguistic informativity? By comparing children's success on the two versions of the present task, we will be able to determine how theory-of-mind abilities are involved in the development of pragmatic abilities.

More broadly, this kind of task can be used to further our understanding of how we calculate and use informativity to accomplish our communicative goals. By manipulating what speakers and/or listeners know about 'context arrays' of the type used in this study, we may be able to learn more about how language users construct models of their conversation partners. For instance, if a speaker learns that a new action, wugging, can be performed only by animals, and not people, how will they take this information into account when designing informative utterances for a partner who either does or doesn't know this? Our understanding of the basic principles of pragmatic inference have expanded greatly from Grice's original formulations, helping us to understand the kinds of evidence that people can take into account when deciding how to help a conversational partner understand what they mean (or how to understand what someone else has said to you. Understanding the origins and development of these abilities will help us clarify why language learning proceeds the way it does, and also give us a much better sense of the fundamental principles that characterize adult language as well.

One intriguing possibility for this work is to understand the patterns of argument dropping that occur not just in adult languages that allow argument drop, but in developing child language. For very young children, it has been consistently observed that children tend to drop subjects more often than they drop objects, producing phrases like *Eat apple* rather than phrases like *Mommy eat*, even though the former is not grammatical in adult English: this behavior tends to occur in both pro-drop and non pro-drop languages. An initial proposal for this pattern was that children simply believe these sentences to be grammatical, until they gather enough information about their native language to learn otherwise (Hyams



& Wexler, 1993). Others have focused on the production limitations of young children: they tend to drop subjects especially when the verb phrase is longer, or has a more challenging metrical (phonological) structure (Bloom, 1990; Gerken, 1991; Valian, 1990). What both of these classes of explanations leave implicit, however, is how the child decides *which* sentences to subject-drop (e.g. versus shortening the object/verb phrase). Subject dropping varies from early development, and there is no sharp cutoff in development between omitting arguments and producing them. A smaller body of work has focused on pragmatic explanations for this variation (Allen, 1997, 2008; Greenfield, 1980; Serratrice, 2007). In these studies, the classification of referents in child speech as accessible or retrievable in discourse can explain much of the variation in both corpora and experimental tasks. These kinds of classifications have also been used to explain the distribution of pronouns vs. full nouns in adult speech, as well as the distribution of subject drop in pro-drop languages (Serratrice 2013.) While corpus analyses in parallel by adult work may be able to determine whether information-theoretic explanations can hold for child speech, the problem of non-linguistic context may be even more important, since children's early speech is so elliptical. Experimental studies like this one can therefore provide an important alternative method for understanding early speech.

As has been the case for understanding the development of word meaning – how children and adults make appropriate generalizations about which words can describe which categories in the world – the principles that are found to explain noun learning may diverge from those which make learning argument structure possible. Nevertheless, shared principles may underlie these processes, and understanding what is both common and different between these learning challenges is an important goal for cognitive science. Expanding our understanding of pragmatic reasoning in adults and children out from referring expressions to argument structure will help us to understand how these important and early-present communicative abilities yield the actual communicative choices that we make in everyday speech.

## FINAL DISCUSSION

In this dissertation, I present a test case for thinking about the fundamental challenges of the connections between higher-level language and cognition. Transitive sentences like “The girl broke the lamp” are one of the simplest structures in language that express a who-did-what-to-whom relation, the basic framework meaning of a sentence. This construction thus provides an ideal way to advance the study of language and cognition, because it requires both coordinating and moving beyond some of the most heavily studied targets of early cognition, namely human agents and simple objects. By 12 months old, as they are just beginning to produce language, infants already reason about human agents: they think about what other people know, about what person’s goal is and how she will achieve it, about how efficient an action is for reaching a particular goal, and about the differences in how humans and inanimate objects move and act (Carpenter, Nagell, & Tomasello, 1998; Gergely, Nadasdy, Csibra, & Biro, 1995; Luo, Kaufman, & Baillargeon, 2009; Phillips & Wellman, 2005; Woodward, 1998). At this point, we know relatively little about how these kinds of cognitive abilities and representations are involved in language. How do children (and adults) coordinate the relationships between entities (a person, a goal object), actions, and language?

I focus on two different types of relationships between language and thought, both somewhat different from the classic cross-linguistic effects of lexical items on perception (cf. Winawer et al., 2007). First, whatever the nature of early (non-linguistic) cognitive representations for events (e.g. those parts of cognition that allow for inferences concerning human actions, causal events, physical and social interactions), they must ultimately be able to be expressed in language: our language-specific representations must refer to our non-linguistic models of the world. Paper 1 establishes that preschoolers’ expectations about the meanings of novel transitive verbs are directly related to their early

nonlinguistic representations of cause. Second, early cognitive representations of other people as social partners – with knowledge and beliefs – appears to underlie some of the choices that young children make about how to understand language or make their own communicative attempts (Carpenter et al., 1998; Nilsen, Graham, Smith, & Chambers, 2008; Stiller, Goodman, & Frank, 2015). Papers 2 and 3 ask how these communicative pressures shape how transitive sentences are expressed, forming the basis of our pragmatic abilities. Paper 2 explores a possible communicative reason for the basic word orders (SOV, SVO) by examining how adults spontaneously order gestures, though ultimately concluding that ordering effects in this task may be due to constraints on production. Paper 3 asks whether adults and children use nonlinguistic context to determine which parts of a transitive sentence are most informative for a listener.

Papers 1, 2, and 3 all address the importance of *argument structure*, the connections between syntax and semantics that establish the basic meaning skeleton of a sentence. Recent approaches to understanding children's inferences, coupled with theories about children's understanding of other people and efficient communication, have made great strides in understanding how children and adults learn names for things and categories of things, and deploy this knowledge to communicate with other people using nouns and noun phrases (cf. Xu & Tenenbaum, 2007). Furthermore, we have learned a great deal about how these kinds of concepts are mapped to language, as well as where those mappings begin to break down (as in the case of mass nouns and non-solid objects, cf. Soja, Carey, & Spelke, 1991). These recent theories of noun learning depend on domain-general inference principles rather than language specific knowledge, so the same principles (e.g. rational communication, size principles for category learning) ought to apply to other aspects of language. This stands in tension with the fact that many have viewed the learning problems for nouns and verbs (and accompanying argument structure) as fundamentally different from one another (Gleitman, 1990; Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005; Pinker, 1994). Among other things, verbs often describe events, which (1) express relationships and (2) are transient in time. Thus there are at least two additional problems: verbs (or more generally the sentences in which they appear) have a different and more elaborated structure than simple kind or entity referents, and there may be additional memory demands, or constraints on what a listener and speaker have in common ground about an event that has occurred in the past. These differences may have impacts for both of the types of communication problems described above, i.e. discovering the semantics of particular sentence structures, and determining the pragmatic felicity of different ways of expressing that structure.

However, if the capacities that allow noun learning to take place as quickly and robustly as it does are truly based on general principles of cognition (whether the representations involved are domain-specific

or not), then these differences may not prevent theory extension. This is true both under theories that posit inference of richly structured representations, as well as theories that propose simple domain-general learning mechanisms (Frank & Goodman, 2014; Smith, 1999). That is, the available evidence and the structure of the concepts to be learned may be very different, but the overarching learning principles may be the same. To determine whether this is the case, we need to better understand the *behavioral signatures* of early argument structure use. Extending our theories of early cognition, language, and communication to cover basic argument structure will thus yield a more general theory of how we form and use our cognitive representations to understand the world – and tell each other about it.

In the three papers of this dissertation, I try to establish what these theory extensions might look like by examining how adults and children solve some of the central problems that are specific to bridging the gap between nonlinguistic abilities and representations, and the basic, but language-specific, structures of simple sentences. Paper 1 focuses on *semantic* connections, showing how one language-specific primitive that affects the syntactic realization of verbs (CAUSE) is related to the corresponding non-linguistic conceptual system: 3- and 4-year-olds who hear novel transitive verbs (but not novel intransitive verbs) expect them to refer to scenes that have the spatiotemporal signatures of causal events, rather than closely matched scenes with the same set of participants and sub-event motions. In addition to clarifying how this particular ‘grounding’ problem is solved at least at one point in development, this kind of work sets the stage for understanding the links between ‘meaning for language’ and ‘meaning for inference’ more generally – transitive sentences can describe other relationships that are not causal, and this kind of paradigm provides some insight into how we can more fully and precisely characterize the expectations that children and adults have about the specific meanings that particular sentence types invoke (complementing existing theories of semantics which mainly rely on speaking judgments about the meanings of sentences rather than comparisons to classes of events.)

Paper 2 moves to the *pragmatic* connections between language and folk theories of communication, by attempting to understand the pressures that produce the basic ordering of agent or subject (S), verb (V), and patient or object (O) of transitive events across the world’s languages. While the paradigm that we use (free and constrained gesture tasks) turned out to have additional unanticipated pressures that complicate our abilities to draw straightforward inferences about this question, this work helps to clarify some of the issues involved in understand how our mental representations of these events get ‘linearized’ in speech for communication. In all of our experiments, we reproduce the initial and puzzling finding that speakers of English (as well as all other languages) prefer to gesture canonical ‘agent-acts-on-inanimate-

object' events in SOV order; one important goal for future work is to try to more fully understand where this bias comes from, and to what degree it reflects how we think about these events for other, non-linguistic tasks. Furthermore, we have asked whether these order effects are also a product of pragmatic pressures, particularly in the case when comprehension is harder and events depart from the more canonical schema (e.g. patients as well as agents are animate.) Although our primary paradigm has confounds that prevent us from drawing conclusions, pilot work in a new symbol-passing task suggests that our expression of basic argument structure is indeed sensitive to the nature of what is being communicated.

Finally, Paper 3 more directly probes the question of what kind of challenges adults and children may face when attempting to communicate about an event to another person. The large literature on referential communication in development has revealed that while adults have sophisticated abilities to refer to objects in the way that can best distinguish them in context, children, at least in their own productions, are sometimes under-informative. However, there is now a growing body of evidence that the beginnings of awareness of informativity emerge around age 3, as children in some tasks and measures seem to be aware of how context affects how the same sentence should be interpreted and whether it does or does not uniquely identify some entity as referent. Paper 3 asks how the challenge changes when you attempt to describe an event rather than an entity, and provides the first evidence that children (aged 5-6) can distinguish between informative and uninformative ways of referring to an event in context. We start with what is potentially the easiest kind of ambiguity one could have in referring expressions for events (i.e. sentences/sentence fragments), namely the number of potential patients and agents. Both adults and children are able to correctly determine whether patients or agents are the most informative; furthermore adults can do this even when the sets of possible participants are large and there is no way to fully identify the event under the communication constraint. In order to better understand these abilities, we will also need to explore how adults and children deal with the other kinds of ambiguities inherent in communicating events, for instance determining *which perspective* on an individual event is to be focused on. However, the present work takes an important step forward in our understanding of informative communication, and particularly its development in younger children.

It is not a new insight to say that language is a special part of human cognition, because it allows us to express propositions from many other areas of cognition or kinds of representations. Over the last several decades, the field has made significant progress toward understanding what kinds of representations and processes can support this ability, identifying a number of ways in which linguistic and non-linguistic

representations are similar and different, and several candidates for overarching theories of cognition that can encompass a wide variety of cognitive abilities including language. In this dissertation, I examine a simple and central kind of linguistic representation, the transitive sentence, which is closely related to a category of canonical events, events that even young infants attend to and make inferences about.

Making the extension from the noun-object kind mapping to syntactically structured language can provide a number of insights. First, there is no single interface between language and cognition. Language both *refers* to representations in other domains, and also *requires* these other abilities for success (e.g. by calculating relevant context sets and evaluating the informativity of possible messages). Second, it reveals critical gaps in our understanding of events. Very preliminary work reveals that adults have shared intuitions about the locations of event boundaries in motion streams (Zacks et al), but a theory of boundaries is no more a theory of events than object segmentation is a complete theory of (object) categorization. The notion of an event may not turn out to be a natural kind for non-linguistic cognition (instead, we may have fundamentally separate ways of expressing interactions in the domains of physics, social interactions, etc.), but a better understanding of this theoretical landscape will be necessary to understanding why human language has the kinds of argument structure that it does. And third, this work points toward the need for higher-level generalizations about the nature of the organization of language. This dissertation explores one simple kind of sentence. In asking what the importance of that particular structure is in its connections to the rest of cognition, it quickly becomes evident that we will need similar explanations of the other basic phenomena at the level of argument structure (e.g. the thematic role hierarchy, the distinction between arguments and modifiers, etc.) Each of these will have a set of interactions with non-linguistic cognition. Are these organizations piecemeal, or do some common set of pressures or featural structures combine to produce the cases we see? We can ask this question both about language (why do we have the constructions and parts of speech that we have?) and non-linguistic cognition (why do we have the core systems/cognitive domains that we do?), and ultimately look for deep parallels among these systems. Can we find evidence of this fundamental organization both in human development and at the neural level? With approaches like those exemplified in this dissertation, maybe we can find out.

## APPENDIX A – ONLINE REPOSITORIES FOR DATA AND STIMULI

Openness in science has a variety of benefits, including reproducibility of analyses and ease of reproduction and extension for other scientists interested in testing similar questions. For this reason, all data (in the form of anonymized coding spreadsheets; we do not have permissions from participants to make video sessions of experiments available), analyses, and stimuli presented in this dissertation are or will be made available online. This information can be accessed in the following places:

### *CHAPTER 2*

Pending: will include anonymized data and analysis documents, plus Matlab code and videos for displaying stimuli to children. Videos are currently available at <http://web.mit.edu/mekline/www/Spatiotemporal/>.

### *CHAPTER 3*

Animation and video stimuli used to present events to participants, code used for online experiments, data sheets, and analysis. Available at:

<http://github.com/mekline/Gesture-Casemarking>

### *CHAPTER 4*

Picture stimuli, code for online adult experiments, data sheets, and analysis available at:

<http://github.com/mekline/SubDrop>

## REFERENCES

- Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development, 67*(2), 635–645.
- Allen, S. (1997). A discourse-pragmatics explanation for the subject-object asymmetry in early null arguments: The principle of informativeness revisited. In A. Sorace, C. Heycock, & R. Shillcock (Eds.), *Proceedings of the GALA '97 Conference on Language Acquisition*. Edinburgh: University of Edinburgh Press.
- Allen, S. (2008). A discourse-pragmatic explanation for argument representation in child Inuktitut. *Linguistics, 48*3–521. <http://doi.org/10.1515/ling.38.3.483>
- Arunachalam, S., Escovar, E., Hansen, M. A., & Waxman, S. R. (2013). Out of sight, but not out of mind: 21-month-olds use syntactic information to learn verbs even in the absence of a corresponding event. *Language and Cognitive Processes, 28*(4), 417–425. <http://doi.org/10.1080/01690965.2011.641744>
- Arunachalam, S., & Waxman, S. R. (2010). Meaning from syntax: Evidence from 2-year-olds. *Cognition, 114*(3), 442–446. <http://doi.org/10.1016/j.cognition.2009.10.015>
- Aylett, M., & Turk, A. (2004). The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence, and Duration in Spontaneous Speech. *Language and Speech, 47*(1), 31–56. <http://doi.org/10.1177/00238309040470010201>



- Baker, C., Saxe, R., & Tenenbaum, J. B. (2011). Bayesian Theory of Mind: Modeling Joint Belief-Desire Attribution. In L. Carlson, C. Hoelscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*.
- Baker, M. (1988). *Incorporation: A Theory of Grammatical Function Changing*. Chicago: Chicago University Press.
- Baldwin, D. A., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, *106*(3), 1382–1407. <http://doi.org/doi:10.1016/j.cognition.2007.07.005>
- Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants Parse Dynamic Action. *Child Development*, *72*(3), 708–717. <http://doi.org/10.1111/1467-8624.00310>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. <http://doi.org/10.1016/j.jml.2012.11.001>
- Bloom, P. (1990). Subjectless Sentences in Child Language. *Linguistic Inquiry*, *21*(4), 491–504.
- Bonawitz, E. B., Ferranti, D., Saxe, R., Gopnik, A., Meltzoff, A. N., Woodward, J., & Schulz, L. E. (2010). Just do it? Investigating the gap between prediction and action in toddlers' causal inferences. *Cognition*, *115*(1), 104–117. <http://doi.org/doi:DOI:10.1016/j.cognition.2009.12.001>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.
- Brown, R. (1958). *Words and Things: An Introduction to Language*. New York, N.Y: Free Press.
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Cantrell, L., & Smith, L. B. (2013). Set size, individuation, and attention to shape. *Cognition*, *126*(2), 258–267. <http://doi.org/10.1016/j.cognition.2012.10.007>

- Christiansen, M. H., & Chater, N. (2009). The myth of language universals and the myth of universal grammar. *Behavioral and Brain Sciences*, 32(05), 452–453.  
<http://doi.org/10.1017/S0140525X09990641>
- Croft, W. (2002). *Typology and universals*. Cambridge: Cambridge University Press.
- Croft, W. (2012). *Verbs: Aspect and argument structure*. Oxford: Oxford University Press.
- Csibra, G. (2003). Teleological and referential understanding of action in infancy. *Phil. Transcripts of the Royal Society of London*, 358, 447–458.
- Cuxac, C., & Sallandre, M.-A. (2007). Iconicity and arbitrariness in French Sign Language: highly iconic structures, degenerated iconicity and grammatic iconicity. In E. Pizzuto, P. Pietrandrea, & Simone (Eds.), *Verbal and Signed languages: Comparing Structure, Constructs and Methodologies* (pp. 13–33.). Berlin: de Gruyter.
- Davies, C., & Katsos, N. (2010). Over-informative children: Production/comprehension asymmetry or tolerance to pragmatic violations? *Lingua*, 120(8), 1956–1972.  
<http://doi.org/10.1016/j.lingua.2010.02.005>
- Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67(3), 547–619.
- Dryer, M. (2005). Order of subject, object, and verb. In *The world atlas of language structures* (pp. 330–333).
- Estes, W. K. (1972). An associative basis for coding and organisation in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 161–190). Washington, D.C.: Winston & Sons.
- Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics*, 21, 95–115.  
<http://doi.org/10.1017/S0142716400001053>

- Fernandes, K. J., Marcus, G. F., Di Nubila, J. A., & Vouloumanos, A. (2006). From semantics to syntax and back again: Argument structure in the third year of life. *Cognition*, *100*(2), B10–B20.  
<http://doi.org/doi:10.1016/j.cognition.2005.08.003>
- Fillmore, C. (1968). The case for case. In E. Bach & R. T. Harms (Eds.), *Universals in Linguistic Theory* (pp. 1–90). New York: Holt Rinehart and Winston.
- Fischer, S. (1975). Influences on word order change in American Sign Language. In C. Li (Ed.), *Word Order and Word Order Change* (pp. 3–25). Austin, TX: University of Texas Press.
- Fisher, C. (1996). Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive Psychology*, *31*(1), 41–81. <http://doi.org/doi:10.1006/cogp.1996.0012>
- Fisher, C., Gertner, Y., Scott, R. M., & Yuan, S. (2010). Syntactic bootstrapping. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 143–149. <http://doi.org/10.1002/wcs.17>
- Fisher, C., Gleitman, L. R., & Gleitman, H. (1991). On the semantic content of subcategorization frames. *Cognitive Psychology*, *23*, 331–392.
- Folli, R., & Harley, H. (2008). Teleology and animacy in external arguments. *Lingua*, *118*(2), 190–202.  
<http://doi.org/10.1016/j.lingua.2007.02.004>
- Frank, M. C., & Goodman, N. D. (2012). Predicting Pragmatic Reasoning in Language Games. *Science*, *336*(6084), 998–998. <http://doi.org/10.1126/science.1218633>
- Frank, M. C., & Goodman, N. D. (2014). Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology*, *75*, 80–96. <http://doi.org/10.1016/j.cogpsych.2014.08.002>
- Futrell, R., Hickey, T., Lee, A., Lim, E., Luchkina, E., & Gibson, E. (2015). Cross-linguistic gestures reflect typological universals: A subject-initial, verb-final bias in speakers of diverse languages. *Cognition*, *136*, 215–221. <http://doi.org/10.1016/j.cognition.2014.11.022>
- Gell-Mann, M., & Ruhlen, M. (2011). The origin and evolution of word order. *Proceedings of the National Academy of Sciences*, *108*(42), 17290–17295. <http://doi.org/10.1073/pnas.1113716108>

- Gergely, G., Nadasdy, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*, 165–193.
- Gerken, L. (1991). The metrical basis for children's subjectless sentences. *Journal of Memory and Language*, *30*(4), 431–451. [http://doi.org/10.1016/0749-596X\(91\)90015-C](http://doi.org/10.1016/0749-596X(91)90015-C)
- Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*(20), 8051–8056. <http://doi.org/10.1073/pnas.1216438110>
- Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A Noisy-Channel Account of Crosslinguistic Word-Order Variation. *Psychological Science*, *24*(7), 1079–1088. <http://doi.org/10.1177/0956797612463705>
- Gillette, J., Gleitman, H., Gleitman, L. R., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*, 135–176.
- Givón, T. (1979). *On understanding grammar*. New York, NY: Academic Press.
- Gleitman, L. R. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*, 3–55.
- Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard Words. *Language Learning and Development*, *1*(1), 23–64. [http://doi.org/10.1207/s15473341lld0101\\_4](http://doi.org/10.1207/s15473341lld0101_4)
- Göksun, T., George, N. R., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). Forces and Motion: How Young Children Understand Causal Events. *Child Development*, *84*(4), 1285–1295. <http://doi.org/10.1111/cdev.12035>
- Goldberg, A. E. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Goldin-Meadow, S., & Feldman, H. (1977). The development of language-like communication without a language model. *Science*, *197*, 401–403.

- Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences*, *105*(27), 9163.
- Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A Theory of Causal Learning in Children: Causal Maps and Bayes Nets. *Psychological Review*, *111*(1), 3–32.
- Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.), *Universals of Language* (pp. 73–113).
- Greenfield, P. M. (1980). Going beyond information theory to explain early word choice: a reply to Roy Pea. *Journal of Child Language*, *7*(01). <http://doi.org/10.1017/S0305000900007145>
- Grice, H. (1975). Logic and Conversation. *Syntax and Semantics*, *3*, 41–58.
- Hale, K., & Keyser, S. (1993). On argument structure and the lexical expression of syntactic relations. In K. Hale & S. Keyser (Eds.), *The view from Building 20: Essays in honor of Sylvain Bromberger*. Oxford: Oxford University Press.
- Hall, M. L., Ahn, Y. D., Mayberry, R. I., & Ferreira, V. S. (2015). Production and comprehension show divergent constituent order preferences: Evidence from elicited pantomime. *Journal of Memory and Language*, *81*, 16–33. <http://doi.org/10.1016/j.jml.2014.12.003>
- Hall, M. L., Ferreira, V. S., & Mayberry, R. I. (2014). Investigating Constituent Order Change With Elicited Pantomime: A Functional Account of SVO Emergence. *Cognitive Science*, n/a–n/a. <http://doi.org/10.1111/cogs.12105>
- Hall, M. L., Mayberry, R. I., & Ferreira, V. S. (2013). Cognitive constraints on constituent order: Evidence from elicited pantomime. *Cognition*, *129*(1), 1–17. <http://doi.org/10.1016/j.cognition.2013.05.004>
- Hamlin, J. K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development*, *26*(1), 30–39. <http://doi.org/10.1016/j.cogdev.2010.09.001>

- Harris, P. L. (2002). What do children learn from testimony? In P. Carruthers, M. Siegal, & S. Stich (Eds.), *Cognitive bases of science*. Cambridge: Cambridge University Press.
- Harris, P. L., & Koenig, M. A. (2006). Trust in Testimony: How Children Learn About Science and Religion. *Child Development*, *77*(3), 505–524. <http://doi.org/10.1111/j.1467-8624.2006.00886.x>
- Henson, R. N. A. (1998). Short-Term Memory for Serial Order: The Start-End Model. *Cognitive Psychology*, *36*(2), 73–137. <http://doi.org/10.1006/cogp.1998.0685>
- Huang, Y. T., & Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental Psychology*, *45*(6), 1723–1739. <http://doi.org/10.1037/a0016704>
- Hyams, N., & Wexler, K. (1993). On the Grammatical Basis of Null Subjects in Child Language. *Linguistic Inquiry*, *24*(3), 421–459.
- Jackendoff, R. (1983). *Semantics and Cognition*. Cambridge, MA: MIT Press.
- Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: The MIT Press.
- Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62. <http://doi.org/10.1016/j.cogpsych.2010.02.002>
- Jaeger, T. F. (2013). Production preferences cannot be understood without reference to communication. *Frontiers in Psychology*, *4*. <http://doi.org/10.3389/fpsyg.2013.00230>
- Kako, E. (2006). The semantics of syntactic frames. *Language & Cognitive Processes*, *21*(5), 1. <http://doi.org/Article>
- Kao, J. T., Levy, R., & Goodman, N. D. (2013). The funny thing about incongruity: A computational model of humor in puns.
- Kao, J. T., Wu, J. Y., Bergen, L., & Goodman, N. D. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, *111*(33), 12002–12007. <http://doi.org/10.1073/pnas.1407479111>

- Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), 25–41.  
[http://doi.org/10.1016/S0010-0277\(03\)00064-7](http://doi.org/10.1016/S0010-0277(03)00064-7)
- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686. <http://doi.org/10.1073/pnas.0707835105>
- Kuhlmeier, V., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological Science*, *14*(5), 402–408.
- Langus, A., & Nespors, M. (2010). Cognitive systems struggling for word order. *Cognitive Psychology*, *60*(4), 291–318. <http://doi.org/10.1016/j.cogpsych.2010.01.004>
- Leslie, A. M. (1984). Spatiotemporal continuity and the perception of causality in infants. *Perception*, *13*, 287–305.
- Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, *25*(3), 265–288. [http://doi.org/doi: DOI: 10.1016/S0010-0277\(87\)80006-9](http://doi.org/doi: DOI: 10.1016/S0010-0277(87)80006-9)
- Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.
- Levin, B., & Rappaport Hovav, M. (2005). *Argument realization*. Cambridge, UK: Cambridge University Press.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.  
<http://doi.org/10.1016/j.cognition.2007.05.006>
- Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*, *106*(50), 21086–21090. <http://doi.org/10.1073/pnas.0907664106>

- Lidz, J., Gleitman, H., & Gleitman, L. (2003). Understanding how input matters: verb learning and the footprint of universal grammar. *Cognition*, 87(3), 151–178. [http://doi.org/doi:10.1016/S0010-0277\(02\)00230-5](http://doi.org/doi:10.1016/S0010-0277(02)00230-5)
- Lillo-Martin, D., & Klima, E. S. (1990). Pointing out differences: ASL pronouns in syntactic theory. In *Theoretical Issues in Sign Language Research, Volume 1: Linguistics* (pp. 191–210). Chicago: University of Chicago Press.
- Luo, Y., Kaufman, L., & Baillargeon, R. (2009). Young infants' reasoning about physical events involving inert and self-propelled objects. *Cognitive Psychology*, 58(4), 441–486. <http://doi.org/10.1016/j.cogpsych.2008.11.001>
- MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, 4. <http://doi.org/10.3389/fpsyg.2013.00226>
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4), 676–703.
- MacKay, D. J. C. (1984). Spoonerisms: The structure of errors in the serial order of speech. In V. A. Fromkin (Ed.), *Speech Errors as Linguistic Evidence*. Walter de Gruyter.
- MacWhinney, B. (1977). Starting points. *Language*, 152–168.
- Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, 126(2), 313–318. <http://doi.org/10.1016/j.cognition.2012.09.010>
- Malchukov, A. L. (2008). Animacy and asymmetries in differential case marking. *Lingua*, 118(2), 203–221. <http://doi.org/10.1016/j.lingua.2007.02.005>
- Markman, E. (1992). Constraints on word learning: speculations about their nature, origins and domain specificity. In M. R. Grunnar & M. Maratsos (Eds.), *Modularity and constraints in language and cognition: the Minnesota Symposium on Child Psychology*. Hillsdale, NJ: Erlbaum.



- Maurits, L., Navarro, D., & Perfors, A. (2010). Why are some word orders more common than others? A uniform information density account. In J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, & A. Culotta (Eds.), *Advances in Neural Information Processing Systems 23* (pp. 1585–1593). Curran Associates, Inc. Retrieved from <http://papers.nips.cc/paper/4085-why-are-some-word-orders-more-common-than-others-a-uniform-information-density-account.pdf>
- McKee, R., & Napier, J. (2002). Interpreting into International Sign Pidgin: An analysis. *Sign Language & Linguistics*, 5(1), 27–54. <http://doi.org/10.1075/sll.5.1.04mck>
- Meir, I., Lifshitz, A., Ilkbasaran, D., & Padden, C. A. (2010). The interaction of animacy and word order in human languages: A study of strategies in a novel communication task. Presented at the 8th international conference on the evolution of language, Utrecht, Germany.
- Meir, I., Sandler, W., Padden, C., & Aronoff, M. (2010). Emerging sign languages. *Oxford Handbook of Deaf Studies, Language, and Education*, 2, 267–280.
- Michotte, A. (1963). *The Perception of Causality*.
- Montague, R. (1970). Universal grammar. *Theoria*, 36(3), 373–398. <http://doi.org/10.1111/j.1755-2567.1970.tb00434.x>
- Morisseau, T., Davies, C., & Matthews, D. (2013). How do 3- and 5-year-olds respond to under- and over-informative utterances? *Journal of Pragmatics*, 59, Part A, 26–39. <http://doi.org/10.1016/j.pragma.2013.03.007>
- Muentener, P., & Carey, S. (2010). Infants' causal representations of state change events. *Cognitive Psychology*, 61(2), 63–86. <http://doi.org/10.1016/j.cogpsych.2010.02.001>
- Nadig, A. S., & Sedivy, J. C. (2002). Evidence of Perspective-Taking Constraints in Children's On-Line Reference Resolution. *Psychological Science*, 13(4), 329–336.
- Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, 17(02), 357–374. <http://doi.org/10.1017/S0305000900013817>

- Naigles, L. (1996). The use of multiple frames in verb learning via syntactic bootstrapping. *Cognition*, 58, 221–251. [http://doi.org/10.1016/0010-0277\(95\)00681-8](http://doi.org/10.1016/0010-0277(95)00681-8)
- Naigles, L. (1998). Developmental changes in the use of structure in verb learning. In C. Rovee-Collier, L. Lipsitt, & H. Haynes (Eds.), *Advances in infancy research* (Vol. 12, pp. 298–318). London: Ablex.
- Naigles, L., & Kako, E. (1993). First Contact in Verb Acquisition: Defining a Role for Syntax. *Child Development*, 64(6), 1665.
- Napoli, D. J., & Sutton-Spence, R. (2014). Order of the major constituents in sign languages: implications for all language. *Frontiers in Psychology*, 5. <http://doi.org/10.3389/fpsyg.2014.00376>
- Nazzi, T., & Gopnik, A. (2000). A shift in children’s use of perceptual and causal cues to categorization. *Developmental Science*, 3(4), 389–396. <http://doi.org/10.1111/1467-7687.00133>
- Nazzi, T., & Gopnik, A. (2001). Linguistic and cognitive abilities in infancy: when does language become a tool for categorization? *Cognition*, 80(3), B11–20.
- Newmeyer, F. (2000). On the reconstruction of “proto-world” word order. In C. Knight, M. Studdert-Kennedy, & J. Hurford (Eds.), *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form* (pp. 372–388). Cambridge: Cambridge University Press.
- Nilsen, E. S., Graham, S. A., Smith, S., & Chambers, C. G. (2008). Preschoolers’ sensitivity to referential ambiguity: evidence for a dissociation between implicit understanding and explicit behavior. *Developmental Science*, 11(4), 556–562. <http://doi.org/10.1111/j.1467-7687.2008.00701.x>
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false belief? *Science*, 308, 255–258.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann Publishers.
- Pesetsky, D. (1995). *Zero Syntax: Experiencers and Cascades*. Cambridge, MA: MIT Press.

- Phillips, A. T., & Wellman, H. M. (2005). Infants' understanding of object-directed action. *Cognition*, *98*, 137–155.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*(9), 3526–3529.  
<http://doi.org/10.1073/pnas.1012551108>
- Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280–291. <http://doi.org/10.1016/j.cognition.2011.10.004>
- Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
- Pinker, S. (1989). *Learnability and Cognition: The acquisition of Argument Structure*. Cambridge, Mass.: MIT Press.
- Pinker, S. (1994). How could a child use verb syntax to learn verb semantics? *Lingua*, *92*, 377–410.
- Rappaport Hovav, M., & Levin, B. (1998). Building verb meanings. In *The projection of arguments: Lexical and compositional factors* (pp. 97–134). Stanford, CA: Center for the Study of Language and Information.
- Resnik, P. (1996). Selectional constraints: An information-theoretic model and its computational realization. *Cognition*, *61*, 127–159.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.
- Schulz, L. E., Standing, H. R., & Bonawitz, E. B. (2008). Word, thought, and deed: The role of object categories in children's inductive inferences and exploratory play. *Developmental Psychology*, *44*(5), 1266–1276. <http://doi.org/10.1037/0012-1649.44.5.1266>
- Scott-Phillips, T. C. (2015). Meaning in animal and human communication. *Animal Cognition*, 1–5.  
<http://doi.org/10.1007/s10071-015-0845-5>

- Scott, R. M., & Fisher, C. (2009). Two-year-olds use distributional cues to interpret transitivity-alternating verbs. *Language and Cognitive Processes*, 24(6), 777 – 803.
- Sedivy, J. (1999). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research*.
- Serratrice, L. (2007). Null and overt subjects at the syntax-discourse interface: Evidence from monolingual and bilingual acquisition. In *The Acquisition of Romance Languages. Selected papers from The Romance Turn II*.
- Shannon, C. E. (1949). Communication in the Presence of Noise. *Proceedings of the IRE*, 37(1), 10–21.  
<http://doi.org/10.1109/JRPROC.1949.232969>
- Snedeker, J., & Gleitman, L. R. (2004). Why it is hard to label our concepts. In S. R. Waxman & Hall (Eds.), *Weaving a Lexicon* (pp. 257–293). Cambridge, MA: MIT Press.
- Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children’s causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science: A Multidisciplinary Journal*, 28(3), 303–333.
- Soja, N. N., Carey, S., & Spelke, E. S. (1991). Ontological categories guide young children’s inductions of word meaning: object terms and substance terms. *Cognition*, 38(2), 179–211.
- Sonnenschein, S., & Whitehurst, G. J. (1984). Developing Referential Communication: A Hierarchy of Skills. *Child Development*, 55(5), 1936. <http://doi.org/10.2307/1129940>
- Spelke, E. S. (1990). Principles of object perception. *Cognitive Science*, 14, 29–56.
- Stiller, A. J., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc Implicature in Preschool Children. *Language Learning and Development*, 11(2), 176–190.  
<http://doi.org/10.1080/15475441.2014.927328>
- Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. *Language Typology and Syntactic Description*, 3, 57–149.

- Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science*, 12(1), 49–100.  
[http://doi.org/doi: DOI: 10.1016/0364-0213\(88\)90008-0](http://doi.org/doi: DOI: 10.1016/0364-0213(88)90008-0)
- Taub, S. (2001). *Language from the Body: Iconicity and Metaphor in American Sign Language*.  
Cambridge: Cambridge University Press.
- Valian, V. (1990). Null subjects: A problem for parameter-setting models of language acquisition.  
*Cognition*, 35(2), 105–122. [http://doi.org/10.1016/0010-0277\(90\)90011-8](http://doi.org/10.1016/0010-0277(90)90011-8)
- Van Son, R. J. J. H., & van Santen, J. P. H. (2005). Duration and spectral balance of intervocalic  
consonants: A case for efficient communication. *Speech Communication*, 47(1–2), 100–123.  
<http://doi.org/10.1016/j.specom.2005.06.005>
- Van Valin, R. D., & LaPolla, R. J. (1997). *Syntax: Structure, Meaning, and Function*. Cambridge:  
Cambridge University Press.
- Weel, J. (2008). *Willow, a Python library for economics*. Retrieved from  
<http://econwillow.sourceforge.net/>
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong  
beliefs in young children's understanding of deception. *Cognition*, 13(1), 103–128.
- Wolff, P. (2008). Dynamics and the perception of causal events. In T. Shipley & J. Zacks (Eds.),  
*Understanding events: How humans see, represent, and act on events* (pp. 555–587). Oxford:  
Oxford University Press.
- Wolff, P., & Song, G. (2003). Models of causation and the semantics of causal verbs. *Cognitive  
Psychology*, 47, 276–332.
- Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, 69, 1–  
34.
- Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, 114,  
245–272. <http://doi.org/10.1037/0033-295X.114.2.245>

- Yuan, S., & Fisher, C. (2009). “Really? She blicked the baby?”: Two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science*, *20*, 619–626. <http://doi.org/10.1111/j.1467-9280.2009.02341.x>
- Yuan, S., Fisher, C., & Snedeker, J. (2012). Counting the nouns: Simple structural cues to verb meaning. *Child Development*, *83*, 1382–1399.
- Zacks, J. M. (2010). The brain’s cutting-room floor: segmentation of narrative cinema. *Frontiers in Human Neuroscience*, *4*. <http://doi.org/10.3389/fnhum.2010.00168>