

## Semantic ambiguity and syntactic bootstrapping: the case of conjoined-subject intransitive sentences

Lucia Pozzan, Lila R. Gleitman & John C. Trueswell

To cite this article: Lucia Pozzan, Lila R. Gleitman & John C. Trueswell (2015): Semantic ambiguity and syntactic bootstrapping: the case of conjoined-subject intransitive sentences, Language Learning and Development, DOI: [10.1080/15475441.2015.1018420](https://doi.org/10.1080/15475441.2015.1018420)

To link to this article: <http://dx.doi.org/10.1080/15475441.2015.1018420>



Published online: 15 Oct 2015.



Submit your article to this journal [↗](#)



Article views: 21



View related articles [↗](#)



View Crossmark data [↗](#)

## Semantic ambiguity and syntactic bootstrapping: the case of conjoined-subject intransitive sentences

Lucia Pozzan, Lila R. Gleitman and John C. Trueswell

Department of Psychology and Institute for Research in Cognitive Science, University of Pennsylvania

### ABSTRACT

When learning verb meanings, learners capitalize on universal linguistic correspondences between syntactic and semantic structure. For instance, upon hearing the transitive sentence “the boy is glorping the girl,” 2-year-olds prefer a two-participant event (e.g., a boy making a girl spin) over two simultaneous one-participant events (a boy and a girl separately spinning). However, 2- and 3-year-olds do not consistently show the opposite preference when hearing conjoined-subject intransitive sentences (“the boy and the girl are glorping”). We hypothesized that such difficulties arise in part from the indeterminacy of the mapping between intransitive syntax and events in the world: a conjoined-subject intransitive sentence can be matched by the one-participant event (if “glorp” means “spin”), both events (“play”), or even the two-participant event (“fight”). A preferential looking study provided evidence for this hypothesis: sentences that plausibly block most non-target interpretations for novel verbs (“the boy and the umbrella are glorping”) eliminated the asymmetric difficulty associated with conjoined-subject intransitives. Thus, while conjoined-subject intransitives clearly pose some special challenges for syntax-guided word learning (“syntactic bootstrapping”) by novices (Gertner & Fisher, 2012), children’s difficulties with this sentence type also reflect expected performance in situations of semantic ambiguity. In discussion, we consider the interacting effects of syntactic- and message-level indeterminacy.

### Introduction

#### *Word learning from scene and syntax*

A crucial question in the cognitive science of language concerns the sources of information and learning mechanisms that account for how children come to say and understand tens of thousands of words before their sixth birthday (e.g., Bloom, 2002), and go on to acquire 50,000 to 100,000 words over the course of a lifetime. On the surface, both the nature of the task and its means of solution look to be straightforward matters of matching up recurrently heard word-length segments (“dog,” “same,” “blue”) with recurrently observed entities, relations, and qualia (dogs, similarity, and blueness).<sup>1</sup> But even leaving aside the provisos and complications from worried philosophers (after all, undetached dog parts are always present when dogs are; Quine 1960), the task looks less and less tractable upon closer inspection. Most word meanings lack simple and invariant physical manifestations. This is most apparent for words that refer to entities absent from the environment at the time of speech, and to the considerable subset of

**CONTACT:** Lucia Pozzan ✉ [lpozzan@sas.upenn.edu](mailto:lpozzan@sas.upenn.edu) 📍 Department of Psychology and Institute for Research in Cognitive Science, University of Pennsylvania, 3401 Walnut Street, Suite 422, Philadelphia, PA 19104.

<sup>1</sup>Notationally, we use “double quotes” for the utterance of a word, ‘single quotes’ for the concept or meaning, and *italics* for the mention of a word.

terms that refer to mind-internal properties (“honest,” “peaceful”), states (“hope,” “know”), and activities (“think,” “intend”), as well as those encoding varying perspectives on the same observable act (“chase”/“flee,” “buy”/“sell”). To infer the meanings of such relational and abstract terms, learners necessarily seek information beyond observation of the immediate physical environment.

According to the theory of syntactic bootstrapping (Gleitman, 1990; Landau & Gleitman, 1985; for precursor theorists, Brown, 1957; C. Chomsky, 1969), much of the required evidence is implicit in the forms of sentences in which particular predicates occur; the licensed structures are a (complex) projection from the predicate semantics. To the extent that this is so, learners can begin to discover the meanings of abstract words via a process that reverse-engineers the meanings from the observed forms. To take a few examples, the number of noun phrases (henceforth, NPs) in a clause tends to line up with the number of participants in a perceived event; the type of complements licensed by a verb reflects mind-internal properties of meaning (e.g., mental and causal verbs license sentential complements); and NP structural position reflects thematic role assignment (Gleitman, 1990; Landau & Gleitman, 1985). Over the years, syntactic bootstrapping as a general approach has enjoyed considerable experimental support (e.g., Fisher, 1996, 2002; Fisher, Gleitman, & Gleitman, 1991; Gertner & Fisher, 2012; Gertner, Fisher, & Eisengart, 2006; Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005; Göksun, Küntay, & Naigles, 2008; Hirsch-Pasek, Golinkoff, & Naigles, 1996; Lee & Naigles, 2008; Lidz, Gleitman, & Gleitman, 2003; Naigles, 1990, 1996; Naigles & Hoff-Ginsberg, 1995; Nappa, Wessell, McEldoon, Gleitman, & Trueswell, 2009; Papafragou, Cassidy, & Gleitman, 2007; Trueswell, Kaufman, Hafri, & Lidz, 2012; Yuan & Fisher, 2009).

Unfortunately, however, not only little children but even adults sometimes fail to oblige investigators by behaving in simple accord with our theories of how structure should be used to recover semantics, raising the question whether the theory itself fails on data, or its predictions are masked by the requirements of some countervailing cognitive principle or task requirement, or whether the investigative process—the choice of stimulus materials, perhaps—is itself at fault. The present paper takes up one important case: the sometime failure of children to map simply between number of arguments and number of thematic roles in intransitive structures. Though the offending case is very particular, it requires resolution lest the explanatory theory be accepted only on its successes, with its failures hastily brushed under the nearest rug.

### ***A failure to parse correctly?***

The crowning jewel instance of structure-guided learning pertains to the one-to-one mapping of thematic roles (constituents of semantic-conceptual structure) onto argument number (constituents of syntactic structure) in the clause. This mapping principle (roughly, the *Theta Criterion* and *Projection Principle* of Chomsky, 1981/1993) is uniformly, though complexly, realized in all known human languages: every predicate argument is expressed as a thematic role in the semantic and syntactic form of a sentence. Accordingly, the earliest (Naigles, 1990) and continuing experimental research in the syntactic bootstrapping tradition has been designed to exhibit young children’s sensitivity to this cue to the meanings of new verbs (e.g., Fisher, 1996; Fisher, Hall, Rakowitz, & Gleitman, 1994; Gertner & Fisher, 2012; Naigles, Gleitman, & Gleitman, 1993; for cross-linguistic documentation, Lidz et al., 2003; Göksun et al., 2008; Trueswell et al., 2012). The idea is that one-argument verbs describe one-participant uncaused events; two-argument verbs describe (or can describe) two-participant, causal, events; and three-argument verbs describe the transfer of entities between parties or places (e.g., “sleep,” “hit,” and “give,” respectively). Of course, ‘syntactic argument’ is itself a quite abstract concept so experimenters have assumed that novice learners identify it, as a first approximation, with the surface feature ‘noun phrase.’ Perhaps, and especially in the largely short and concrete-content sentences of caretaker-child speech, the learner could count NPs as rough proxies for counting arguments. We say such a procedure would be “rough” because of

complex NPs (“the brother of John”), conjoined NPs (“John and Bill”), and adjunct phrases (“in the morning”), so there are bound to be sentences with more NPs than arguments; and all languages allow for NP surface deletion under certain conditions, so there are sometimes too few NPs (“Pick up your blocks”; “Sue just ate”; “Jill told Jack to eat”).

### **The test case**

In her classic first studies of the learning effects of implicit syntactic analysis, Naigles (1990) showed that mere babies (average age 25 months) will infer aspects of a new verb’s meaning, namely, its argument structure, by listening to a sentence in the presence of an ambiguously interpretable scene. As one instance, a video is shown in which Duck pushes Bunny down, thus causing Bunny to squat, whilst both these characters simultaneously twirl one of their arms. While watching this scene, half the children heard a novel verb (“glorp”) embedded in a transitive sentence (1), and the other half heard the same verb in an intransitive sentence (2).

- (1) *The duck is glorping the bunny.*
- (2) *The duck and the bunny are glorping.*

The participants were then shown “disambiguated scenes,” either of Duck pushing down Bunny (with no arm-twirling) or of the two characters independently arm-twirling (but no pushing down), while hearing a syntactically uninformative verbal prompt: “Where’s glorping now? Find glorping.” Those infants who had heard (1) when they were first introduced to “glorp” looked longer at the two-participant event in the subsequent “disambiguated” scene, while those who had initially heard (2) displayed the opposite tendency. Knowledge of something like Chomsky’s theta criterion and projection principle thus seems to be guiding and constraining children’s hypotheses regarding the meaning of a new verb: during training, an ambiguous world is apparently resolved by unambiguous syntax; then this fixation of meaning persists to determine the domain of semantic applicability of the word (“glorp”).

Since publication of Naigles (1990), numerous researchers have replicated her findings in this and related paradigms for transitive sentences (Arunachalam & Waxman, 2010; Bavin & Growcott, 2000; Gertner et al., 2006; Gertner & Fisher, 2012; Hirsh-Pasek et al., 1996 for 27- to 30-month-olds only; Kidd, Bavin, & Rhodes, 2001; Naigles & Kako, 1993; Noble, Rowland, & Pine, 2011; Noble, Theakston, & Lieven, 2010; Yuan & Fisher, 2009). However, a much more complex and inconsistent picture emerges when the structure under investigation is as in (2), namely the conjoined-subject intransitive (henceforth CI). It is indeed the case that a number of studies replicate the effect of syntax observed in the Naigles study, namely that hearing a CI sentence such as (2) reduces the preference to look at the two-participant event (Arunachalam & Waxman, 2010; Kidd et al., 2001).

However, and crucially motivating the experimentation we will present here, it is not the case that hearing a CI sentence consistently results in the expected preference to map this sentence to the target one-participant event. In some studies the child participants did exhibit such a preference (Naigles, 1990; Noble, Rowland, & Pine, 2011 [with 40- and 50-month-olds]; Noble, Theakston, & Lieven, 2010 [with 51 month-olds]), but just as many other experiments failed to replicate this preferential behavior (Arunachalam & Waxman, 2010; Naigles & Kako, 1993; Noble et al., 2011 [with 27- and 31-month-olds]; Noble et al., 2010 [with 30-month-olds and 40-month-olds]). Instead, preference for the two depicted events did not differ significantly.

We believe there are two plausible reasons why children present inconsistent preferences in response to CI sentence types. First, children, especially young children who are just breaking into the sentential syntax of their native language, ought to show an inability to build the correct structure for input strings intended to denote the CI sentence structure; owing to universal biases and incomplete language-specific knowledge, young children should fail to distinguish between CI and transitive sentence types and thus show a preference to map CI sentences onto two-participant causal events just as they do for transitive sentences. As discussed below in *A potential failure to*

find the parse of the sentence, Gertner and Fisher (2012) recently tested and found experimental support for this hypothesis. Immediately following this section we discuss why this account, although likely true, cannot provide a full explanation of children's (and adults') interpretive errors with CI sentences, motivating a second potential reason for interpretive inconsistencies (see *A potential misreading of the intended message*), namely the possibility that the child has failed, instead or in addition, to parse the scene itself for the intended message. The experimental findings below explore this account in detail.

### ***A potential failure to find the parse of the sentence***

As Gertner and Fisher (2012) note, any scheme that identifies NP's with argument positions is going to run into trouble with conjoined-nominal sentences. To understand that "the duck and the bunny" exemplifies a single argument rather than two, one must know a number of linguistic facts particular to English. The first is that "and" is a connective effecting a coordination of two (or more) phrasal categories (here, NPs) which together constitute a higher (plural) NP. The second is that "are" implies plural agreement with the sentence subject. If these language-specific facts are ignored (e.g., the child instead thinks the connective "and" and the auxiliary "are" are inflectional particles), then this sequence will be analyzed as two separate NPs. Under a count-the-NPs heuristic, this mis-parse will lead to the false inference that there are two arguments in the structure, which, in turn, must refer to a two-participant event, QED sentence (2) and its new predicate ("glorp") would be systematically misunderstood such that these young learners will (just as for sentence (1)) look longest, or at least as long, at the two-participant event.

As Gertner and Fisher reasoned, if this chain of inference is what underlies the behavior of children who do not match (2) with one-participant events, they face yet one further interpretive decision: which of these two presumed arguments is the subject (hence the actor or causal agent) and which is the object (hence, the patient)? Learners might be at sea, deciding at random; but they might not: they might systematically rely on the tendency in languages of the world for agents to precede patients serially (for this claim and documentation, Bates & MacWhinney, 1982; Bever, 1970; Slobin & Bever, 1982; cf., Dittmar, Abbot-Smith, Lieven, & Tomasello, 2008). The bottom line is that (2) would now be interpreted as an agent-patient-verb structure: "the duck the bunny is glorping" (meaning "the duck is glorping the bunny").

Gertner and Fisher (2012) tested this hypothesis in young learners of English. Using the Naigles (1990) preferential-looking paradigm, they provided 21-month-old children with a video scene depicting a two-participant event (e.g., a girl rocking a boy on a chair) and a video scene depicting two independent one-participant events (e.g., a boy and a girl next to each other, both bouncing on fitness balls). The children failed to distinguish between these structures and consistently preferred to look at the video depicting the two-participant events both when they heard transitive sentences such as (3) and when they heard CI sentences like (4), thus replicating effects in the literature that we have already described, and that can be accounted for on a count-the-surface-NP strategy under the subject-object (SO) structural reading. But what about sentence (5) below? Here, crucially, the sentence would mismatch both of the observed scenes because the first (subject) NP is "the boy" and the boy is not causing anything; rather, he is the patient of the rocking event. The result, in accord with Gertner and Fisher's logic, was that for this kind of sentence looks to the two-participant scene dropped dramatically because in this case the two-participant scene (girl-rocking-boy) was not a potential candidate for the agent-patient parse (boy-rocking-girl).

- (3) *The girl is glorping the boy.*
- (4) *The girl and the boy are glorping.*
- (5) *The boy and the girl are glorping.*

Summarizing, these results are plausibly interpreted to indicate that these infants were (a) counting NPs as a proxy for counting argument positions, (b) interpreting the serially first NP as

subject/agent of the action and the second NP as the patient; and (c) matching both transitive and intransitive two-NP's sentences with two-participant scenes when the surface order (girl-boy) of the NP's was compatible with the agent-patient (girl = agent, boy = patient) structure of the depicted event.

### ***Shadows in a beautiful picture: some residual questions***

The Gertner and Fisher (2012) findings and explanations certainly have moved us closer to understanding why CI sentences might create so much interpretive trouble for young listeners. But some details and collateral findings from other laboratories are puzzling if this kind of explanation is the whole story. One caveat is that, in addition to the assumptions made so far, to accept the explanation that children are interpreting NP-NP sequences as agent-patient structures, we would have to assume that children at this age are ignoring cross-cutting morphological cues (“is” vs. “are”; “and” as a conjunction) that contradict their presumed analysis. In addition, an implicit assumption of this analysis is that the unknown lexical element (“glorp”), appearing in final position in (4)-(5), will be analyzed as the verb (predicate) element by default. After all, it is a universally valid principle across the languages of the world, not just English, that a verb-like element will surface in the clause. In (4)-(5), it is in final position. If children at this age have even weakly generalized, based on past experience, that verbs surface medially in English (i.e., that English is SVO), for the present (NP NP V) analysis to hold, they would have to be ignoring the word-order cue of English that verbs canonically precede rather than follow objects. Also puzzling are recent findings showing that much older listeners in similar experimental circumstances—listeners who surely know that English does not permit the order agent-patient-verb—also do not consistently match CI sentences with the target one-participant event scene (for children, see Noble et al., 2011; Noble et al., 2010; for adults, see Pulverman, Capote, Hughson, Garber, & Sorrell, 2012; Sheline, Waxman, & Arunachalam, 2013). “Something else” must be accounting for why these advanced speaker-listeners, who understand the basic canonical facts of English grammar, nevertheless fail in this one condition of selective-looking experiments.

A final property of listeners' behavior in this paradigm is even harder to explain as being the product of misparsing or deriving only from a partial syntactic analysis. It is the observation that children's difficulties with intransitive sentences persist when there is a plural NP subject as in (6) (see Noble et al., 2010):

(6) *The animals are glorping.*

Here the count-the-surface-NP strategy cannot be the basis of error, for in this case universal and language-specific syntactic constraints alike should lead the learner to assign a one-argument structure to the surface string.

### ***A potential misreading of the Intended message***

There is a different account of what goes wrong when young children fail to map simply between intransitive structures and depicted one-participant events. It could be that the fault lies not in their parse but rather in their failure to interpret the events depicted in the accompanying video scenarios in the way experimenters intended. That is, even when children successfully parse a CI sentence as intransitive, there very well could exist interpretations of a two-participant causal-event scene that would plausibly map onto the intransitive sentence structure. This explanation, for which we will present evidence in the body of this paper, is not in opposition to the explanation from Gertner and Fisher, but rather is by hypothesis another major source of variance in how infants and toddlers line up heard sentences with the scenes to which they pertain. After all, even assuming that the speaker's sentence is causally related to the scene in view—that the speaker *by his/her words* means to describe *that scene*—there is in principle no

end of representations in terms of how this particular word-to-world pairing might be satisfied. We need not belabor this issue which has been discussed in the philosophical and psychological literature just about forever (in the present learning context, see Chomsky, 1957). Many different messages can be appropriate and plausible for the same observed scene. It might very well be that it is not always so easy to see the world as others are seeing it.

We propose that word learners might not pair conjoined-subject and plural-subject intransitive sentences with one-participant event videos because of such indeterminacies in how sentences refer to and describe events. At least sometimes, the problem is the message as reconstructed by the listener, rather than the sentence form. Specifically, the intransitive sentence may plausibly be describing one-participant events that are co-present within the two-participant event videos (for a related remark, see Arunachalam & Waxman, 2010). For instance, a video of a boy causing a girl to twirl is also a video of *a boy and a girl smiling, playing, or having fun*. Therefore the observing, listening child may devote attention equally to the experimenter-designated one-participant event scene (the target event) *and* to the experimenter-designated two-participant event scene when hearing “the duck and the bunny are glorping.” The animals are having fun in both depictions, so looking behavior may well be 50/50. In contrast, congruence of scene with a transitive sentence is more selective, for it applies predominantly when a cause-effect or doer-undergoer relation is depicted in the observed world. Then we expect a systematic preference for the experimenter-designated two-participant scene when the heard sentence is transitive.

Summarizing, there is a potential asymmetry between one- and two-participant sentence structures, which might in turn be responsible for young children’s consistent performance with the former, but not with the latter. Such an account predicts the finding that two-NP two-argument structures elicit selective looking at (what the experimenter takes to be) two-participant scenes, but more variable responsiveness to two-NP one-argument structures (to repeat: because both depictions conceivably represent one-participant sub-events co-present in the two-participant scenes).

### **Experimental prospectus**

To test this depiction-indeterminacy hypothesis, we manipulated the animacy of the second NP in transitive and CI sentences ((3) and (4), repeated here as (7) and (8), compared to (9) and (10)), along with the animacy of the corresponding entities in the videos.

- (7) *The girl is glorping the boy*
- (8) *The girl and the boy are glorping*
- (9) *The girl is glorping the box*
- (10) *The girl and the box are glorping*

The motive was to reduce the number of possible interpretations of the target two-participant events. To be sure, animate and inanimate entities can both undergo the same process simultaneously without interacting with each other: a girl and a box can both roll, fall, or get wet together, and so forth. However, the idea of a girl and a box playing together or having fun together, interacting in any meaningful way (let alone a box smiling!), is much less plausible, hardly even a candidate for interpretation, *pace* the socially interacting dishes and coffee pots in Disney’s *Beauty and the Beast*.<sup>2</sup> The predicted upshot was that few participants would be likely to interpret the two-participant video interaction of

<sup>2</sup>These intuitions were supported by the results of a separate web-based norming study conducted with native adult speakers. Participants [N=52] viewed two-participant event videos with Animate-Animate and Animate-Inanimate participants and heard a novel verb either within a transitive or CI sentence, like (7)-(10). Participants then wrote what the verb meant. Sub-event interpretations such as “have fun”, “smile” or “play” were frequent if such scenes are described by CI sentences like “the boy and the girl are glorping” (43%). Sub-event descriptions of this type were far less common (21%,  $\chi^2(1) = 10.66, p < .001$ ) when the same event involved an animate entity acting on an inanimate object while hearing the corresponding CI sentence (e.g., “the boy and the umbrella are glorping”). In this case the modal response was to leave the response blank, indicating that subjects cannot provide an intransitive predicate that applies to both entities (21% for animate-animate CI vs. 35% for animate-inanimate ones,  $\chi^2(1) = 4.04, p = .04$ ).

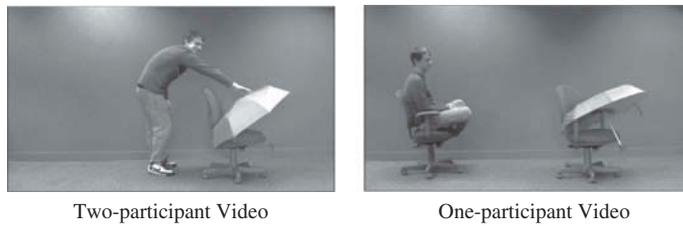


Figure 1. Example of Animate-Inanimate video stimuli.

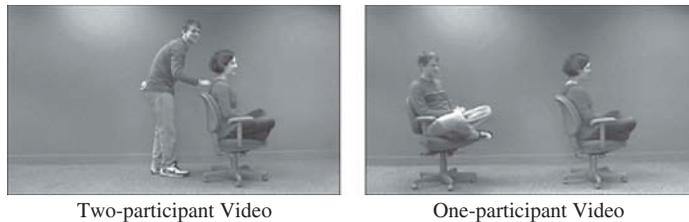


Figure 2. Example of Animate-Animate video stimuli.

umbrella and boy in Figure 1 as a “smiling”, “playing”, or “having fun” event and in turn as a plausible meaning for the novel verb in (8), though they might do so with the interaction scene of boy and girl in Figure 2.

If this prediction is supported, it would suggest that children may be correctly interpreting the conjoined-NP constructions but are less successful at pairing them with the target scenes in the ways intended by the investigating scientists. Notice that, in our experimental materials (see (9)-(10)), the inanimate NP is always serially second, thus in the “patient” position. This is for two reasons. The first is that this is the problematic case in Gertner and Fisher’s (2012) study in that the second NP in the CI sentence corresponds to the patient in the two-participant scene. The second is that agent/patient relations prototypically (though not necessarily) involve an animate agent acting on an inanimate patient (e.g., Croft, 1991; Langacker, 1987). Indeed, if children are incorrectly parsing the CI sentence type as an agent-patient-verb construction, the expected outcome is opposite of what our account predicts: children might be *even more* tempted to interpret CI sentences as referring to the two-participant event scene when the first NP is animate and the second inanimate, as this semantic pattern further bolsters an agent-patient analysis and a prototypical causal relation.

Finally, we will be testing both a group of 2-year-olds and a group of 3- and 4-year-olds. The reason to test a group of older children was to diminish the plausibility that failure to parse correctly can be the entire cause of the difficulties seen with CI syntax. Children of this relatively advanced age, in fact, freely utter CIs (Lukyanenko & Fisher, 2012; Tager-Flusberg, de Villiers, & Hakuta, 1982) but still display difficulties with CI and plural-subject intransitive sentences in tests of comprehension (Noble et al., 2010, 2011). We therefore expect to find that our predictions regarding animacy will also hold in older children. In addition, comprehension data were collected for a small sample of adult native speakers because two recent studies have found a similar lack of consistent target-like performance with CI sentences in adult participants (Pulverman et al., 2012; Sheline et al., 2013). This simply cannot be because these adults do not understand even the rudiments of English grammar in which sentences are canonically SVO, “are” carries an agreement morpheme, and “and” is a sentential connective.

## Experiment

We present a preferential looking experiment conducted with 2- to 4-year-olds, along with a small sample of adult controls. Like related studies in the literature, critical trials involved participants viewing two videos simultaneously side-by-side, one depicting an entity causing another entity to do something (a two-participant event scene) and another video in which the same two entities are now carrying out the corresponding action independently (one-participant event scene). See [Figures 1](#) and [2](#) for examples of the video pairs used in the Animate-Inanimate and the Animate-Animate conditions, respectively.

A group of 2- to 4-year-olds watched these videos and heard a novel verb embedded in either a transitive or a CI frame; we refer to this group as the **experimental group**. A second group of 2- to 4-year-olds watched the same videos and heard the novel verbs presented within a neutral syntactic context (“Look! I see glorping! See? Find glorping!”); we refer to this group as the **baseline group**.

In the Animate-Animate condition, we used videos similar to those used in the previous literature, in which two animate entities served as the event participants (i.e., a boy and a girl, [Figure 2](#)). The aim was to investigate whether difficulties pairing CI sentences with target one-participant event scenes can also be observed in older children, who presumably have the necessary language-specific knowledge to parse these sentences correctly. The Animate-Inanimate condition was identical to the Animate-Animate one except that the depicted events involved a person and an inanimate object ([Figure 1](#)). Animacy was manipulated between subjects for both the experimental and the control groups.

Differently from related studies in the literature, we (1) tested the effect of syntax on children’s preference for the two-participant event scene using a within-subjects rather than a between-subjects design, (2) presented four experimental items per condition (eight total) rather than one or two, (3) constructed our materials so that the same general event was depicted in the two- and one-participant video pairs (i.e., both videos depicted a chair-spinning event in [Figures 1](#) and [2](#)), and (4) recorded eye-movements with an eye-tracker instead of video-cameras, which allowed us to collect more data per time sample (30 Hz vs. 60Hz) and did not require human coding.

## Method

### Participants

A total of 149 children participated in the study. Data from seventeen children were excluded, due to program malfunctioning (N=1), high eye-track loss (N=9), language background (N=5), or experimenter error (N=2). Data from 132 children (Age Range = 24-59 months, Mean Age = 40 months, SD = 10 months) were analyzed. Children were randomly assigned to the animacy conditions. Children in the two animacy conditions did not differ in terms of age (Animate-Animate: Mean Age = 39, SD = 10; Animate-Inanimate: Mean Age: 41, SD = 10;  $t(131) = .99, p = .32$ ) or gender make-up (27 females in the Animate-Animate condition and 33 females in the Animate-Inanimate condition). Of these 132 children, the experimental group comprised 88 participants (Age Range: 25–56 months, Mean Age = 39 months, SD = 10), 45 (17 females) in the Animate-Animate and 43 (20 females) in the Animate-Inanimate conditions. The baseline group comprised the remaining 44 participants (Age Range: 24–59 months, Mean Age = 42 months; SD = 10), 22 in the Animate-Animate (10 females) and 22 in the Animate-Inanimate (13 females) conditions. In addition, 16 students (9 females) at the University of Pennsylvania participated (8 in each animacy condition). They received course credit for their participation. They were all native speakers of English.

## Apparatus

A Tobii T120 eyetracker was used to record eye movements. The T120 sampled eye position at 60 Hz and displayed the stimuli via its built-in 17" monitor set to 1024x768 pixels (43.2 cm viewing size). The spatial resolution of the T120 is approximately 0.5 degrees visual angle. Stimulus display was controlled by E-Prime experiment software version 2.0.8.79 in the Windows XP operating system, running on a Dell Precision M4400 laptop with 3.48 GB RAM, a 2.66 GHz dual processor, and an NVIDIA Quadro FX 770M card with 512MB video memory. The laptop was disconnected from the internet to increase timing accuracy. Audio for the stimuli was presented through two speakers positioned on each side of the T120 monitor.

## Materials and design

Sixteen experimental video pairs (eight for each animacy condition) were created. Each video pair consisted of a two-participant and one-participant version of the same event. For example, if the general event was "spinning on a chair," the two-participant event video depicted character A making character B spin on a chair, while the one-participant event video depicted A and B simultaneously spinning on their own chairs. Two types of events were depicted in the videos: *change of place* events that happened through a medium (e.g., spinning on a chair) or *change of state* events (e.g., lighting up).<sup>3</sup> Video pairs in the Animate-Animate condition (e.g., Figure 2) consisted of a video in which a girl acted on a boy (and vice versa, in half of the trials) and a video in which a girl and a boy simultaneously underwent the same process. A corresponding set of video pairs depicting the same events were prepared for the Animate-Inanimate condition (e.g., Figure 1); these consisted of a video in which a girl (or a boy, in half of the trials) acted on an inanimate object and a video in which a girl and the object simultaneously underwent the same process (see Table 1 for the complete list of video pairs).

Two practice trials for each animacy condition were created in the same manner as above, except that the events were labeled with known verbs. In the experimental version of the study, one of the practice sentences was transitive (e.g., "The boy is pushing the girl"), and the other was a CI (e.g., "The boy and the girl are falling").

An Animate-Animate stimuli list and a corresponding Animate-Inanimate stimuli list were created. Each list consisted of two practice trials, followed by eight experimental trials. In the experimental version of the study, half of the experimental trials used transitive syntax whereas the other half used CI syntax. In the baseline condition, verbs were presented within neutral syntax (e.g., "I see glorping!"; "I see nizzing!"). Within each of the conditions, half of the items had the two-participant event video on the left and half had the two-participant event video on the right. Experimental trial order was the same for both lists and consisted of a single fixed random order with transitive and CI items; the presentation of transitive and CI items was blocked. A second version of each experimental list was then created by changing the transitive items to CI and the CI to transitive, resulting in four stimuli lists in total. Finally, six additional

<sup>3</sup>Two anonymous reviewers noted that some of the one-participant *change of state* events do not seem to be good candidate referents for intransitive syntax descriptions, and that such events would be better described by the use of a passive construction in English (e.g., "the girl and the boy are getting sprayed"). This, however, depends on the particular English verb that one uses to translate the novel verb with (e.g., if the verb were "shower" rather than "spray", a description using intransitive syntax would be more appropriate than a description using passive syntax). More generally, the crucial point is that *change of state* and *change of place* events tend to pair with unaccusative and ergative verbs in English. In intransitive sentences, the subject of these predicates is associated with a theme/experiencer theta role analysis, making CI sentences suitable descriptions of one-participant, noncausal, events. In order to investigate whether the type of event (i.e., *change of state* vs. *change of place*) depicted in the videos had a significant effect on participants' looking preferences, additional analyses were conducted that tested for effects of event type and any interaction with the other factors in this experiment. Specifically, it was found that overall looking times for the experimental conditions (Transitive, CI) were unaffected by event-type and event-type did not interact with other factors for both the 3- to-4-year olds (all  $p$ 's > .05) and the 2-year-olds (all  $p$ 's > .05). Event-type did not influence overall looking patterns in adults either. We thus did not include event-type as a factor in analyses reported below.

**Table 1.** Descriptions of the two- and one-participant event scenes, along with arbitrarily associated novel verbs.

Verb	Two-participant event	One-participant event
<i>Blick</i>	The girl illuminates the boy/box using a big lamp	The girl and the boy/box become illuminated (lamps not visible)
<i>Dax</i>	The boy pours water on the girl/pillow	The boy and the girl/pillow become wet (pitcher not visible)
<i>Heaf</i>	The boy pushes the girl/pillow and makes her/it slide down a slide	The boy and the girl/pillow slide down a slide
<i>Glorp</i>	The girl pushes the boy/ball and makes him/it roll on a skateboard	The girl and the boy/ball roll on skateboards
<i>Loodge</i>	The girl sprays pink silly string on the boy/ball.	The girl and the boy/ball get covered in pink silly string (spray bottle not visible)
<i>Niz</i>	The boy makes the girl/umbrella spin on a chair	The boy and the girl/umbrella spin on chairs
<i>Pim</i>	The boy pours talc powder on the girl/umbrella	The boy and the girl/umbrella get 'showered' with talc powder (talc powder bottles not visible)
<i>Plock</i>	The girl pushes the boy/box and makes him/it swing	The girl and the boy/box swing

lists (four for the experimental version of the study, and two for the baseline condition) were created in which the order of the experimental trials and the left-right positions of the videos were reversed. For the baseline condition, everything was the same as in the experimental version, with the exception of the syntax manipulation.

### Procedure

Participants were randomly assigned to one of the lists and tested individually in a quiet, moderately lit room. They sat comfortably in front of the eye tracker and participated in a calibration procedure followed by the experimental procedure.

A standard 5-point calibration procedure was used, as implemented in Tobii Studio version 2.1.14. During the procedure, a red cross-hair moved around the screen, arriving at each of five locations (the center and each of the four corners). Immediately after the calibration routine, the experimenter judged the calibration quality by examining the Studio calibration plot for the number of calibration points for which eyegaze data were collected and the spread of data around those points. If the calibration data did not meet the default criteria of the Studio software or if more than one point was missing data, the calibration routine was repeated.

The experimental procedure was closely based on Gertner and Fisher (2012) and is exemplified in Table 2. It began with two practice trials, followed by eight experimental trials. Each trial began with a character-presentation phase (A1, see Table 2), in which a picture of one of the characters was presented alone on one side of the screen and labeled twice while the other side remained blank. The second character was then introduced in the same way on the other side of the screen (A2). Participants' attention was then brought to the center of the screen by having a smiley face appear accompanied by the sound of a doorbell (B). Participants needed to fixate the smiley face in order for the next phase to begin. Next, they were asked to identify each of the characters (C1, C2): both characters appeared simultaneously on the screen, and children were prompted to find each of the two characters in turn. Next, the events were introduced. Each event was viewed by itself on one screen and accompanied by neutral language while the other side of the screen remained blank (D1, D2). Participants' attention was again brought to the center of the screen by having a smiley face appear together with the sound of a doorbell (B). Participants needed to fixate the smiley face in order for the next phase to begin. Next, both videos were played simultaneously on the screen and one of them would be described once by the experimental sentence (E1). The onset of the experimental sentence was preceded by two seconds in which the child saw both videos accompanied by neutral language (e.g., "Hey look!"). This provided the child with enough time to check out both videos while they were both present on the screen. Each video lasted between 6 and 7 seconds. The videos would then freeze on the last frame and would be described by the same sentence type in past tense (E2). Participants' attention was again brought to the center of the

**Table 2.** Stimulus trial structure.

Phase	Duration	Screen set-up		Sentence/sound
<i>A1. Character Presentation</i>	4 sec.		Umbrella	Hey, look! There's an umbrella. Look at the umbrella.
<i>A2. Character Presentation</i>	4 sec.	Boy		Hey, look! There's a boy. Look at the boy.
<i>B. Fixation</i>			Smiley Face	<i>Doorbell sound</i>
<i>C1. Character Identification</i>	4 sec.	Umbrella	Boy	Find the umbrella. Where's the umbrella?
<i>C2. Character Identification</i>	4 sec.	Umbrella	Boy	Find the boy. Where's the boy?
<i>D1. Event Presentation</i>	6.5 sec.	Video 1		Look here! Watch this!
<i>D2. Event Presentation</i>	6.5 sec.		Video 2	Look over here! Watch this!
<i>B. Fixation</i>			Smiley Face	<i>Doorbell sound</i>
<i>E1. Event Description</i>	6.5 sec	Video 1	Video 2	Hey, watch! The boy is nizzing the umbrella. See?
<i>E2. Event Description</i>	6.5 sec.	Last Frame of Video 1	Last Frame of Video 2	Oh wow! The boy nizzed the umbrella. Let's watch it again!
<i>B. Fixation</i>			Smiley Face	<i>Doorbell sound</i>
<i>E3. Event Description</i>	6.5 sec.	Video 1	Video 2	The boy is nizzing the umbrella. Find nizzing!

screen by having a smiley face appear on the screen together with the sound of a doorbell (B). Finally, the videos were played simultaneously a second time and described by the experimental sentence (E3). At the end of the trial, a picture of a running 'Hello Kitty' appeared on the screen while the eye-gaze data was being saved.

### **Analysis of eye movement data**

Trials were excluded from further analysis if track loss occurred during more than 40% of the presentation time of each experimental video. Participants (Children:  $N=9$ ; Adults:  $N=1$ ) were excluded if track loss resulted in less than four items in total (two per syntactic condition in the experimental version of the study). In total, 17% of children's trials and 10% of adults' trials were removed due to track loss; once these items were removed, track loss amounted on average to 8% for children and 10% for adults.

For each trial, two areas of interest (AOIs) were defined corresponding to the one-participant and the two-participant event videos. Participants were considered to be looking at a video if the eye position was within the pixel coordinates of the video in question. Pixel coordinates outside these two AOIs, as well as track loss, were treated as looks to elsewhere. For children, at a typical viewing distance of 20 inches, each AOI subtended 18 degrees visual angle horizontally and 12 degrees visual angle vertically; for adults, at a typical viewing distance of 26 inches, each AOI subtended 15 degrees visual angle horizontally and 9 degrees visual angle vertically. The neighboring edges of these two AOIs were separated by 0.5 degrees visual angle horizontally. This distance allowed for accurate discrimination of looks to each AOI, given the eyetracker's 0.5 degree resolution.

In order to best compare our results with previous findings in the literature, the presentation of the results is divided into two sections. First, we present a coarse-grained measure of participants' preference for the two-participant event scene (see "Overall Looking Times"), as indicated by the amount of time spent looking at the two-participant event scene compared to the amount of time spent looking at the one-participant event scene (henceforth, two-participant event preference score), from the onset of the experimental sentence until the end of the video presentations (E1 and E3). Second, we present a more fine-grained analysis of participants' preference for the two-participant event video as it unfolds over time (see "Fine-grained Eye Movement Analysis"), from the onset of the experimental sentence until 4800 ms after it. This time frame was selected to allow equal sampling of time course information for the two video presentations. This analysis will give us insight into the developing interpretation as each portion of the sentence is heard.

Inferential statistics on overall looking times are presented as the results of linear mixed-effects models with the maximal random-effects structures justified by the models. Time-course analyses

are based on growth curve models, a multilevel regression technique that has been developed to model change across time and has been applied to both longitudinal data and eye-tracking data within the visual world paradigm (for a detailed summary of the method, see Mirman, 2014; Mirman, Dixon, & Magnuson, 2008).

Both set of analyses were performed using R software (version 3.1.3) with R adds-on lme4 and lmerTest. The latter package, which provides type 3 and type 1 F tests for fixed effects, and Likelihood Ratio Test (LRT) tests for random effects, was used to calculate p-values for the maximal models; the denominator degrees of freedom for these tests were calculated using the Satterthwaite approximation.

For simplicity, we graph looking preferences in terms of proportion difference scores but perform inferential statistical analyses on empirical logit (elogit, henceforth) transformed data (Barr, 2008). In particular, the transformed looking preference was calculated by taking, for each child/item combination, the natural log of: the time spent looking at the two-participant video plus 0.5 all divided by the time spent looking at the one-participant video plus 0.5.

## Results and discussion

Here we report how children's two-participant event preference scores change as a function of the syntax in which novel verbs are presented (Transitive vs. CI syntax vs. Baseline), the animacy of the two event participants (Animate-Animate vs. Animate-Inanimate), and children's age. For both animacy conditions, we expected to replicate the finding in the literature that children in this age range discriminate between transitive and CI syntax, in that they spend more time looking at two-participant events when presented with transitive than with CI syntax. In addition, for Animate-Animate contexts, we expected to replicate past observations that children prefer the two-participant events as the referent for transitive sentences but might not consistently prefer the one-participant event scene as the referent for CI sentences; that is, their preferences might not differ from participants who heard a neutral sentence like "I see glorping!" as was used in the baseline condition. Crucially, however, we expected children exposed to Animate-Inanimate contexts to be more likely than children in Animate-Animate ones to pair CI sentences with one-participant events, since sub-event components of Animate-Inanimate two-participant events are less likely to be categorized as one-participant events compatible with CI descriptions (e.g., "play," "have fun," "smile").

### Overall looking times

Figures 3 and 4 present average two-participant event preference scores as a function of Syntax (Transitive, CI) and Age (in months) for children in the Animate-Animate (Figure 3) and Animate-Inanimate (Figure 4) experimental (nonbaseline) conditions. Preference scores were calculated averaging across a broad time window, from the onset of the experimental sentence until the end of the video presentation, collapsing across the first (E1) and the second video presentations (E3). Positive values indicate a preference to look at the two-participant more than at the one-participant event videos, while negative values indicate the opposite preference. The gray circles indicate Transitive sentences, and the black squares indicate CI sentences.

A linear mixed-effects model was applied to the (elogit) preference data, and included fixed effects of Age (a continuous variable in months), Syntax (Transitive, CI) and Animacy (Animate-Animate, Animate-Inanimate), plus all interactions. The results of this model, together with information about its random-effects structure appear in Table 3. As can be seen in Table 3, an overall reliable effect of Syntax was observed, such that the children's preference to look at the two-participant-event video was greater when they heard Transitive sentences than when they heard CI sentences. Thus, children as a group were sensitive to syntactic information in the expected direction; they were more likely to map the Transitive than the CI sentences onto two-participant events. However, this sensitivity to

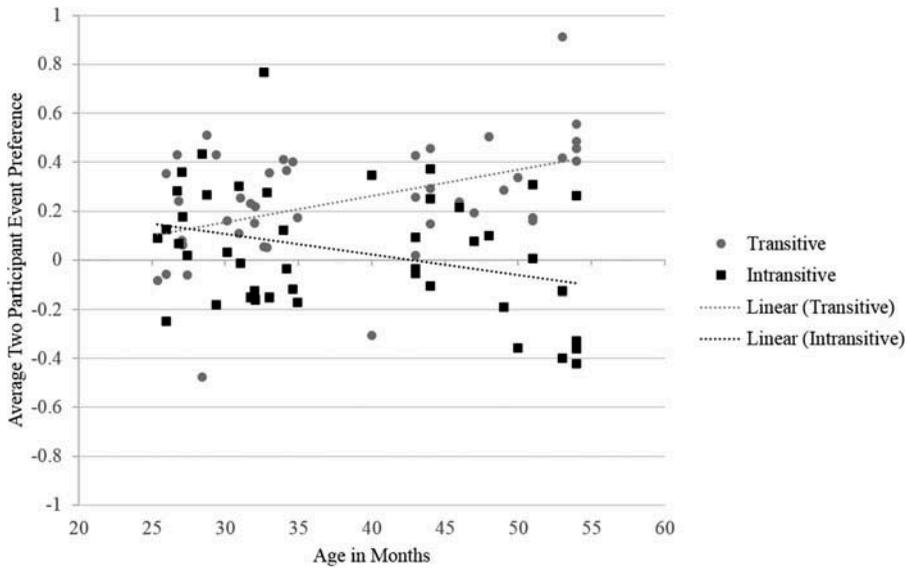


Figure 3. Two-participant overall looking preferences as a function of syntax and age for Animate-Animate contexts.

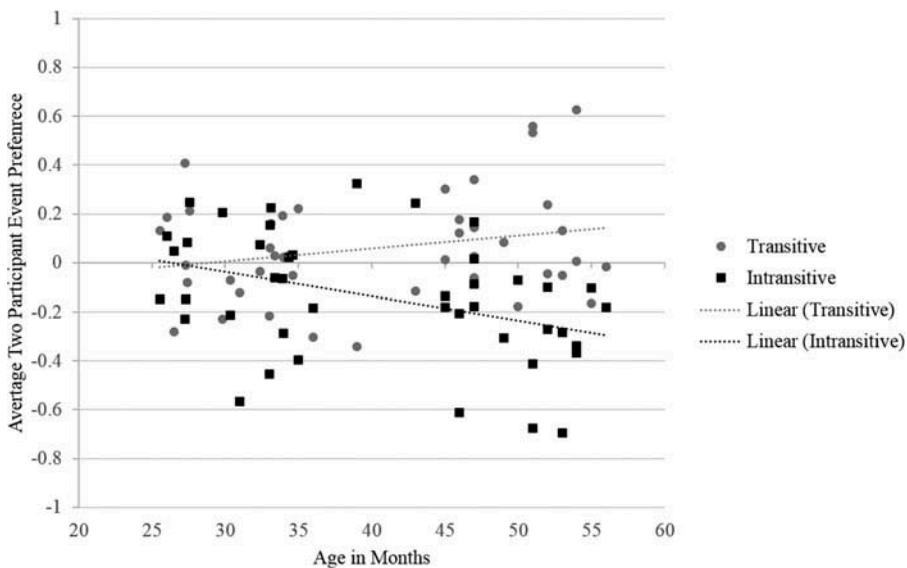


Figure 4. Two-participant overall looking preferences as a function of syntax and age for Animate-Inanimate contexts.

Syntax was found to interact with Age (Age x Syntax), such that the effect of Syntax increased with Age. To explore this interaction further, we conducted additional analyses (not reported in the table) for the two different sentence types separately. Consistent with the shape of the interaction, we found that two-participant event preference scores increased with increasing Age for Transitive sentences ( $Estimate = .02$ ,  $SE = .01$ ,  $t(10) = 2.28$ ,  $p < .05$ ) and decreased with increasing Age for CI sentences ( $Estimate = -.02$ ,  $SE = .01$ ,  $t(12) = -2.84$ ,  $p < .05$ ). Finally, the primary model (Table 3) also revealed that two-participant event preference scores were overall higher in Animate-Animate than Animate-Inanimate contexts (effect of Animacy) but this effect did not interact with Age (Age x Animacy) nor Syntax (Animacy x Syntax). Indeed, the separate analyses performed on each sentence type revealed

**Table 3.** Linear mixed effect model of (elogit) two-participant overall looking preferences including effects of age, syntax and animacy for experimental conditions (transitive and CI sentence) only.

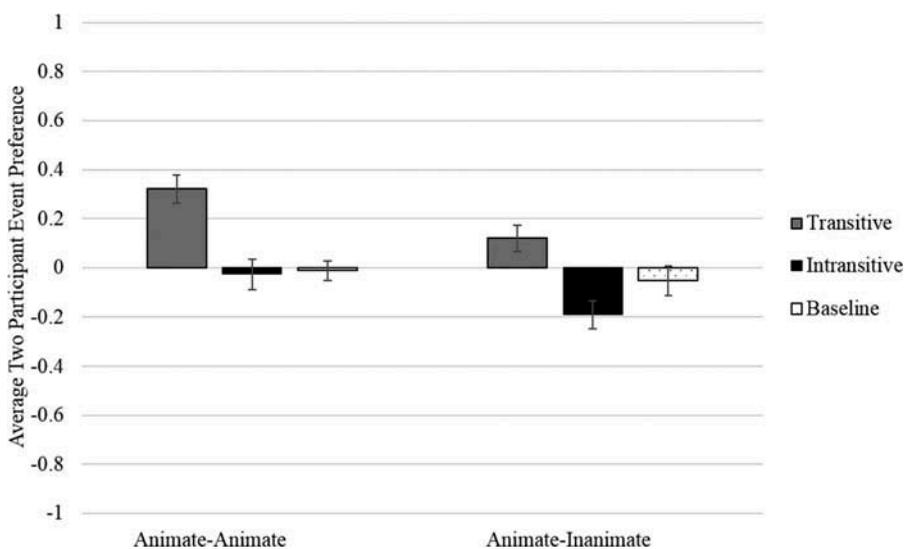
Fixed effects	Estimate	S.E.	df	t	p-value	Significance
Intercept	0.139	0.11	8	1.27	0.238	
Age (mo.)	-0.001	0.00	86	-0.17	0.864	
Syntax	0.510	0.13	15	4.00	0.001	*
Animacy (AA vs. AI)	0.479	0.09	17	5.25	<0.001	*
Age x Syntax	0.047	0.01	87	4.49	<0.001	*
Age x Animacy	0.007	0.01	86	0.80	0.428	
Animacy x Syntax	0.159	0.20	86	0.78	0.439	
Age x Animacy x Syntax	0.007	0.02	87	0.35	0.727	

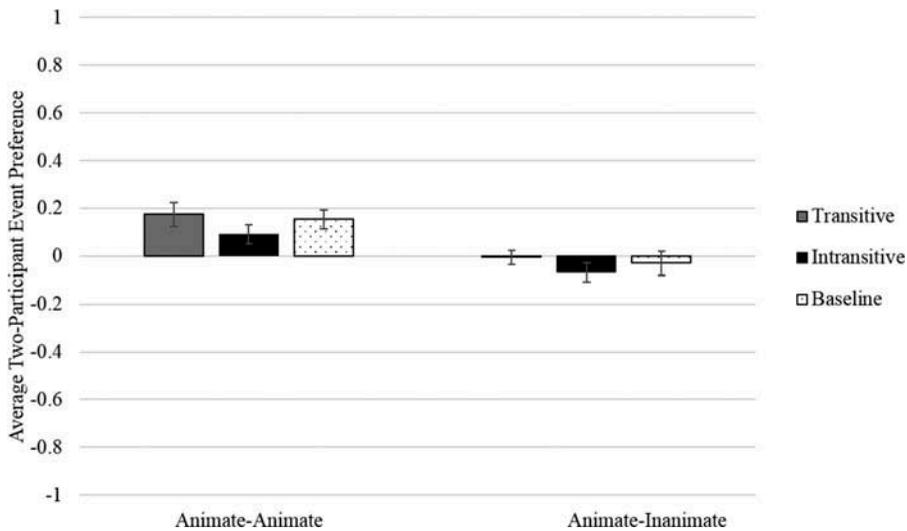
Notes. AA= Animate-Animate; AI = Animate-Inanimate; CI = Conjoined Subject Intransitive; Random effects structure included intercept and effect of Syntax for Subjects plus intercept and effects of Animacy and Syntax for Items. R-code used with lmerTest:  $\text{Elogit} \sim \text{Age} * \text{Animacy} * \text{Syntax} + (1 + \text{Syntax} | \text{Subject}) + (1 + \text{Animacy} + \text{Syntax} | \text{Item})$ . Degrees of freedom (df) calculated using the Satterthwaite approximation.

that this positive effect of Animacy held for both Transitive ( $\text{Estimate} = .57, SE = .16, t(16) = 3.51, p < .01$ ) and CI sentences ( $\text{Estimate} = .38, SE = .13, t(29) = 2.96, p < .01$ ). No other effects or interactions were significant.

To examine further how children's looking preferences were modulated by the syntax of the sentence and the age of the child, we split our subjects into two age groups reflecting the age clusters seen in Figures 3 and 4: the first group (henceforth 2-year-olds) ranged in age from 24 to 36 months; the second group (henceforth 3- to 4-year-olds) ranged in age from 37 to 56 months. Figure 5 presents the average two-participant event preference for 3- to 4-year-olds whereas Figure 6 presents the corresponding results for 2-year-olds. In addition, for both age groups, average results from the children in the baseline condition are also plotted for comparison (i.e., the results from those children who were provided uninformative syntactic evidence: "Look! I see glorping! Find glorping!"). We consider first the results from the 3- to 4-year-olds and then consider the results from the 2-year-olds.

**Three- to 4-year-olds.** As can be seen in Figure 5, 3- to 4-year-olds' preference to look at the two-participant event over the one-participant event followed the predicted patterns. For Animate-Animate contexts, Transitive sentences elicited a preference to look more at the two-participant

**Figure 5.** Three- to-four-year-olds: two-participant overall looking preferences as a function of animacy and syntax.



**Figure 6.** Two-year-olds: two-participant overall looking preferences as a function of animacy and syntax.

event, whereas CI sentences elicited no preference and had a value similar to the neutral syntax Baseline condition. These observations were supported by the results of a linear mixed-effects model applied to these data (Table 4), which included a three-level fixed effect of Syntax (Baseline, Transitive, CI) in which the Baseline was treated as the reference (comparison) condition. It was found that Transitive sentences generated a significantly greater two-participant looking preference than the Baseline condition, whereas CI sentences did not differ from the Baseline. For the Animate-Inanimate contexts, the same mixed-effects model (Table 5) revealed that Transitive sentences were reliably greater than the Baseline whereas CI sentences trended toward being lower than Baseline. Separate versions of both of these models were also run without the Baseline condition (not reported in the tables) thereby permitting a test of the contrast between Transitive and CI sentences; as expected, this contrast was significant in both Animate-Animate (Estimate =  $-.89$ , SE =  $.19$ ,  $t(136) = -4.56$ ,  $p < .01$ ) and Animate-Inanimate (Estimate =  $-.77$ , SE =  $.14$ ,  $t(136) = -5.62$ ,  $p < .01$ ) contexts, replicating past work showing that children are sensitive to this syntactic difference.

Finally, it should be noted that, as seen in Figure 3, 3- to 4-year-old children in the Baseline conditions showed little or no preference for one of the two video scenes in either Animacy condition, as indicated by the non-significant intercepts in Tables 4 and 5. This suggests that children in this age range do not, in the absence of syntactic evidence, have a preference to look at one video over the other.

Thus the findings from this age group are consistent with the past literature in that they suggest that children consistently match transitive syntax with two-participant events but do not consistently pick either event as the referent for CI sentences. These results are also consistent with our experimental hypothesis that children in this age range possess the relevant language-specific

**Table 4.** Linear mixed effect model of (elogit) two-participant overall looking preferences in Animate-Animate contexts including effect syntax (baseline, transitive, CI) for 3- to 4-year-olds.

Fixed effects	Estimate	S.E.	df	t	p-value	Significance
Intercept	-0.039	0.23	9	-0.17	0.867	
Syntax (Baseline vs. Transitive)	0.855	0.30	9	2.82	0.019	*
Syntax (Baseline vs. CI)	-0.036	0.22	9	-0.16	0.876	

Notes. CI = Conjoined Subject Intransitive; Random effects structure included intercept Subjects plus intercept and effect of Syntax for Items. R-code used with lmerTest:  $\text{Elogit} \sim \text{Syntax} + (1|\text{Subject}) + (1+\text{Syntax}|\text{Item})$ . Degrees of freedom (df) calculated using the Satterthwaite approximation.

**Table 5.** Linear mixed effect model of (elogit) two-participant overall looking preferences in Animate-Inanimate contexts including effect syntax (baseline, transitive, CI) for 3- to 4-year-olds.

Fixed effects	Estimate	S.E.	df	t	p-value	Significance
Intercept	-0.105	0.12	10	-0.86	0.410	
Syntax (Baseline vs. Transitive)	0.395	0.16	8	2.39	0.045	*
Syntax (Baseline vs. CI)	-0.373	0.17	9	-2.17	0.060	

Notes. CI = Conjoined Subject Intransitive; Random effects structure included intercept for Subjects plus intercept and effect of Syntax for Items. R-code used with lmerTest:  $\text{Elogit} \sim \text{Syntax} + (1|\text{Subject}) + (1+\text{Syntax}|\text{Item})$ . Degrees of freedom (df) calculated using the Satterthwaite approximation.

knowledge to prefer one-participant events as the referent of CI sentences, once various one-participant events that are co-present with two-participant event scenes (e.g., “play,” “smile,” “have fun,” etc.) are rendered unlikely candidates for novel predicate meanings, as done in the Animate-Inanimate condition.

**Two-year-olds.** As can be seen in Figure 6, the 2-year-olds in the experimental (nonbaseline) conditions provided average looking patterns that resembled those of the older children, in that two-participant event preference scores were numerically higher for Transitive than for CI sentences, in both Animate-Animate and Animate-Inanimate contexts. However, linear mixed-effects models on these data (Tables 6 and 7) revealed no reliable differences from Baseline preferences for either Transitive or CI sentences in either the Animate-Animate or Animate-Inanimate conditions. Moreover, separate tests directly comparing the two experimental conditions (Transitive vs. CI) were also not significant (both  $p$ 's > .10).

It should be noted that, as seen in Figure 6, Baseline preferences may be affecting the results for two year-olds. In the Animate-Animate baseline condition, children showed a marginally significant preference for two-participant events ( $\text{Estimate} = .45$ ,  $\text{SE} = .20$ ,  $t(7) = 2.28$ ,  $p = .06$ ), while in the Animate-Inanimate baseline condition, children did not show a preference for either scene ( $\text{Estimate} = -.08$ ,  $\text{SE} = .18$ ,  $t(8) = -0.46$ ,  $p = .65$ ). Nevertheless, the observed null effect of syntactic structure is surprising given previous studies in the literature, which have shown that children in this age range are able to discriminate between Transitive and CI sentences (Araunachalam & Waxman, 2010; Gertner & Fisher, 2012 [for 25-month-olds]; Kidd et al., 2001). It is possible that reliable differences between the two syntactic conditions were obscured by baseline preferences or by the use of a fairly

**Table 6.** Linear mixed effect model of (elogit) two-participant overall looking preferences in Animate-Inanimate contexts including effect syntax (baseline, transitive, CI) for 2-year-olds.

Fixed effects	Estimate	S.E.	df	t	p-value	Significance
Intercept	-0.080	0.18	8	-0.44	0.672	
Syntax (Baseline vs. Transitive)	0.049	0.15	7	0.32	0.759	
Syntax (Baseline vs. CI)	-0.092	0.15	14	-0.60	0.558	

Notes. CI = Conjoined Subject Intransitive; Random effects structure included intercept for Subjects plus intercept and effect of Syntax for Items. R-code used with lmerTest:  $\text{Elogit} \sim \text{Syntax} + (1|\text{Subject}) + (1+\text{Syntax}|\text{Item})$ . Degrees of freedom (df) calculated using the Satterthwaite approximation.

**Table 7.** Linear mixed effect model of (elogit) two-participant overall looking preferences in Animate-Animate contexts including effect syntax (baseline, transitive, CI) for 2-year-olds.

Fixed effects	Estimate	S.E.	df	t	p-value	Significance
Intercept	0.450	0.19	8	2.33	0.049	*
Syntax (Baseline vs. Transitive)	0.051	0.18	4	0.28	0.797	
Syntax (Baseline vs. CI)	-0.193	0.23	0	-0.78	0.800	

Notes. CI = Conjoined Subject Intransitive; Random effects structure included intercept for Subjects plus intercept and effect of Syntax for Items. R-code used with lmerTest:  $\text{Elogit} \sim \text{Syntax} + (1|\text{Subject}) + (1+\text{Syntax}|\text{Item})$ . Degrees of freedom (df) calculated using the Satterthwaite approximation.

broad time window. We explore this latter possibility later in the section “Fine-grained Eye-Movement Analysis.”

**Adults.** Similar analyses were carried out to investigate adult speakers’ comprehension of these sentences. As can be seen in Figure 7, two-participant event preference scores were, as expected, higher for Transitive than for CI sentences, regardless of Animacy. This is confirmed by the statistical analyses. The model for this analysis included the fixed effects of Animacy, Syntax and their interaction. The random-effect structure included by-subject and by-item random intercepts, together with by-subject random slopes for the main effect of Syntax and by-item random slopes for the main effects of Syntax, Animacy, and their interaction. There was a significant effect of Syntax ( $Estimate = -4.60$ ,  $SE = .62$ ,  $t(14) = -7.39$ ,  $p < .01$ ), no effect of Animacy ( $Estimate = -.15$ ,  $SE = .47$ ,  $p = .66$ ), and no interaction between Syntax and Animacy ( $Estimate = -.77$ ,  $SE = 1.16$ ,  $p = .52$ ).

To further investigate whether adults not only distinguish between the two syntactic structure types, but also prefer different events as the referents of Transitive and CI sentences, we then explored whether adults’ showed a significant preference (or dis-preference) for the two-participant scene in each of the conditions separately. Due to the fact that baseline effects were weak in the child participants (not significant in the older group of children, only present in the Animate-Animate condition in the younger group), we did not collect baseline data for adults; adults’ two-participant event preference scores for each of the syntactic and animacy conditions separately were compared with chance (zero) instead. As can be seen in Figure 7, adults’ two-participant event preference scores were significantly above zero for Transitive sentences and significantly below zero for CI sentences, regardless of Animacy (Transitive Sentences: Animate-Animate:  $Estimate = 2.17$ ,  $SE = .46$ ,  $t(8) = 4.68$ ,  $p < .01$ ; Animate-Inanimate:  $Estimate = 2.69$ ,  $SE = .54$ ,  $t(6) = 5.00$ ,  $p < .01$ ; CI Sentences: Animate-Animate:  $Estimate = -2.02$ ,  $SE = .70$ ,  $t(6) = -2.89$ ,  $p = .02$ ; Animate-Inanimate:  $-2.33$ ,  $SE = .44$ ,  $t(6) = -5.28$ ,  $p < .01$ ).

Adult participants’ looking patterns thus behave in accord with the classic intuition (Naigles, 1990) that transitive and CI structures tend to refer to different types of events in the world (for a discussion of why children and adults might differ in this respect, see the “General Discussion” section).

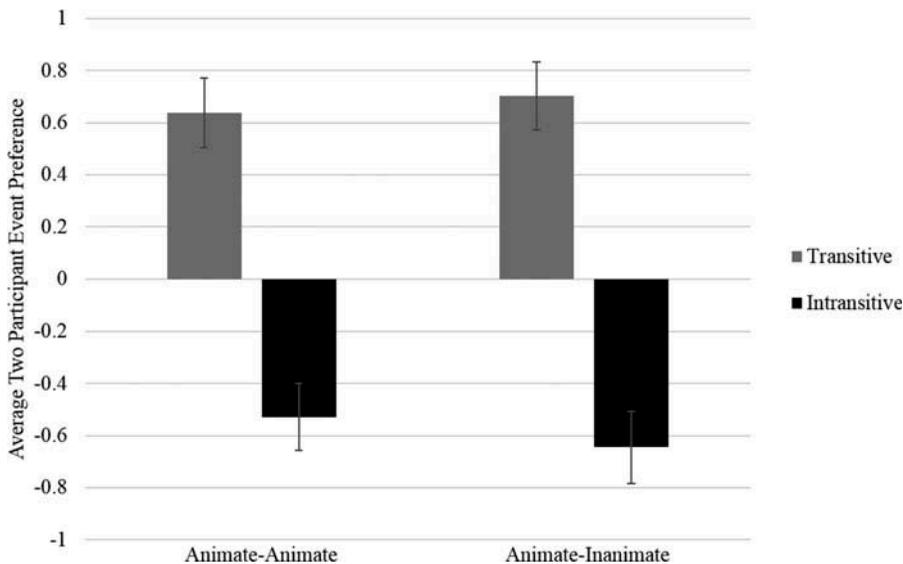


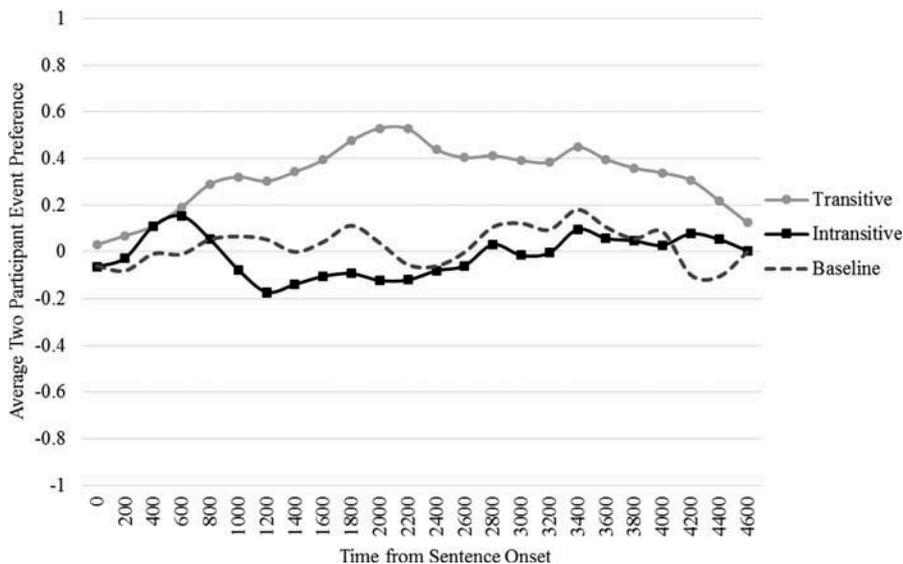
Figure 7. Adult participants: two-participant overall looking preferences as a function of animacy and syntax.

### Fine-grained eye-movement analysis

Figures 8, 9, 10, and 11 plot children's average two-participant event preference scores over time from the onset of the experimental sentence up to 4800 ms after it, collapsing across the first and the second presentations of the experimental video pairs (E1 and E3). Eye movements were sampled every 16 ms, but for the purposes of this analysis two-participant event preference scores were computed collapsing the data into 200 ms time-bins.

Growth curve analyses (Mirman, 2014) were used to analyze the time course data for Animate-Animate and Animate-Inanimate conditions separately. In all these analyses, first-order (linear), second-order (quadratic) and third-order (cubic) orthogonal polynomials were used to model the gaze patterns across time. Each model also included the three-level factor of Syntax (Baseline, Transitive, CI), together with the interaction between each syntactic contrast (Baseline vs. Transitive; Baseline vs. CI) and the Time factors. As explained in Mirman et al. (2008), an effect of a given factor on the intercept (i.e., a main effect) reflects an overall difference in the two curves, shifting one upward or downward overall; an interaction between the fixed factor and the linear time term reflects a difference in the angle of two curves (i.e., a difference in the overall rise or fall in looking preference); an interaction with the quadratic term reflects a difference in the symmetric rise and fall around a central point of inflection; finally, an interaction with the cubic term reflects a difference in the steepness of the curve around two inflection points.

The R add-on package *lmerTest* was used to calculate p-values for the maximal models, which always included the fixed effects of Syntax, Time and their interactions. The random-effect structure included by subject random intercepts, together with by-subject random slopes for the Time factors. In addition, in order to ensure that the inclusion of the orthogonal polynomials used to model change across time did not result in data overfitting, we conducted a parallel set of analyses comparing the p-values obtained using the *lmerTest* package with those obtained using the likelihood ratio test. These two methods yielded analogous results, in that whenever *lmerTest* indicated that a given fixed effect was significant, the inclusion of that fixed effect resulted in a significant improvement to the model fit; estimates and p-values for the models reported below are based on the analyses performed using the *lmerTest* package.



**Figure 8.** Animate-Animate condition: 3- to 4-year-olds' two-participant overall looking preferences as a function of syntax and time.

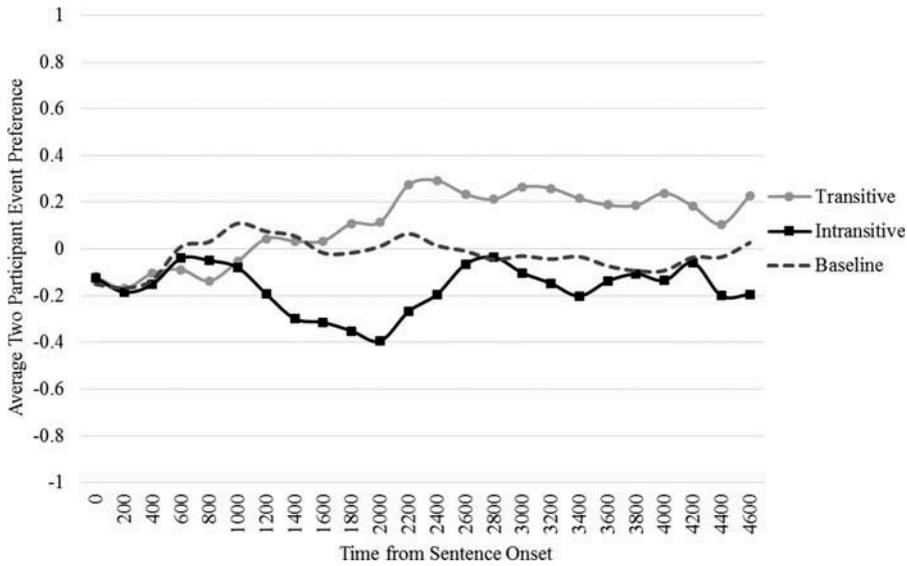


Figure 9. Animate-Inanimate condition: 3- to 4-year-olds' two-participant overall looking preferences as a function of syntax and time.

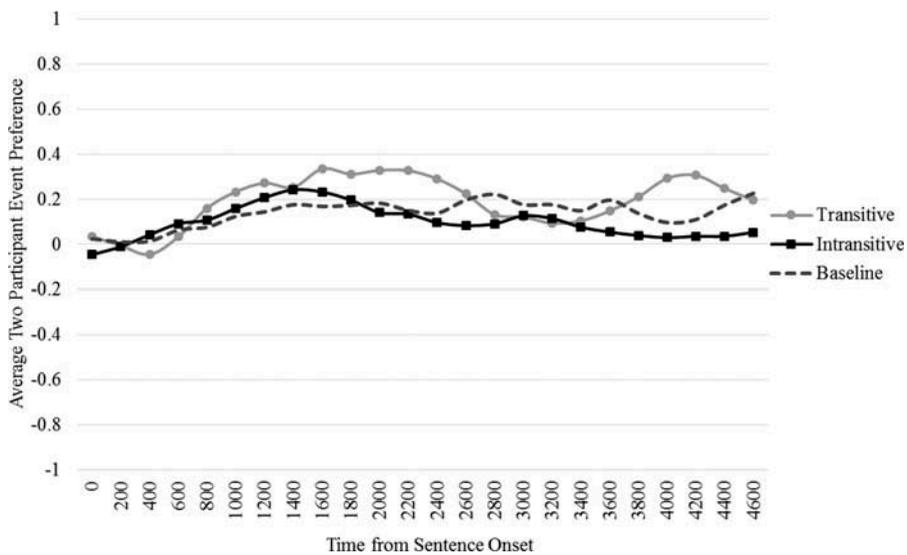
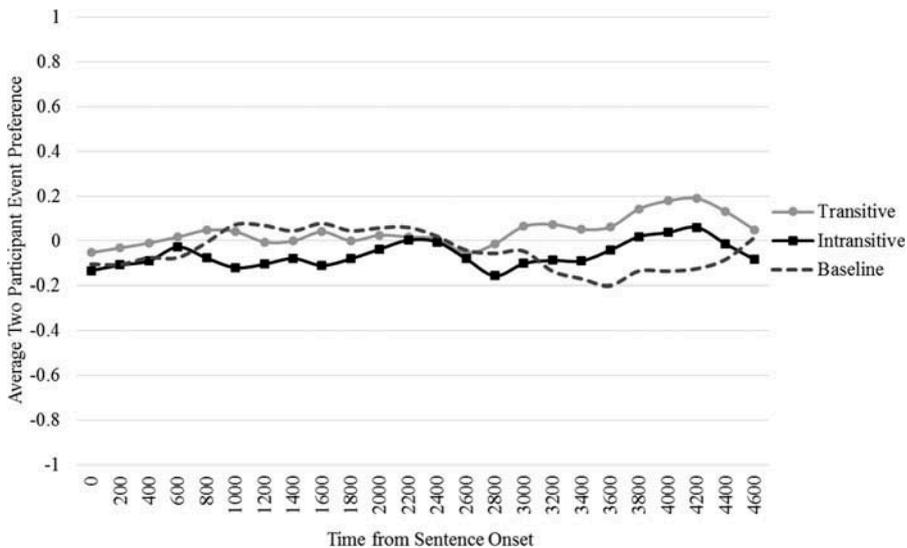


Figure 10. Animate-Animate condition: 2-year-olds' two-participant overall looking preferences as a function of syntax and time.

*Three- to 4-year-olds.* As can be seen in Figure 8, the time-course results for 3- to 4-year-old children in the Animate-Animate Condition produced changes in looking preferences similar to what was observed in the overall looking times reported above: hearing a Transitive sentence triggered increased looks to the two-participant event, whereas hearing a CI sentence in the same context triggered no systematic changes in looking preferences and showed a pattern similar to the Baseline (neutral syntax) condition. The growth curve analysis for this condition (Table 8) supported these observations. As can be seen in the table, the syntactic contrast of Baseline vs. Transitive had a reliable positive effect on the intercept, indicating overall greater preference to look at the two-participant event when hearing Transitive syntax as compared to hearing Baseline neutral syntax. In addition, this same syntactic contrast (Baseline vs. Transitive) interacted with the quadratic time



**Figure 11.** Animate-Inanimate condition: 2-year-olds' two-participant overall looking preferences as a function of syntax and time.

term, indicating greater steepness in the central inflection for the Transitive condition. In contrast, the syntactic contrast of Baseline versus CI yielded no significant effects or interactions; the intercept term (reflecting looking preferences in the Baseline condition) was also not significantly different from zero. Taken together, all these null effects suggest there was little or no change in looking preferences for these latter two conditions across time. Finally, we conducted an additional growth-curve analysis (not shown in Table 8) that directly compared the two experimental conditions (Transitive vs. CI). This analysis confirmed that the two experimental conditions were indeed different from each other; there was a reliable effect of the syntactic contrast (Transitive vs. CI) on the intercept and the quadratic time term (both  $p$ 's <.01).

The time-course results for 3- to 4-year-old children in the Animate-Inanimate condition (Figure 9) produced a pattern consistent with the overall looking times reported above, but here temporal dynamics of the looking preferences provide additional information. Specifically, Transitive sentences deviate from the neutral Baseline at a later point in time than CI sentences, with the CI

**Table 8.** Growth curve analysis: linear mixed effect model of (elogit) two-participant overall looking preferences in Animate-Animate contexts including effects of time and syntax (baseline, transitive, CI) for 3- to 4-year-olds.

Fixed effects	Estimate	S.E.	df	t	p-value	Significance
Intercept	0.003	0.15	39	0.02	0.983	
Time1	0.233	0.49	51	0.47	0.639	
Time2	-0.091	0.60	43	-0.15	0.881	
Time3	-0.260	0.45	49	-0.58	0.568	
Syntax (Baseline vs. Transitive)	1.000	0.19	39	5.30	<0.001	*
Syntax (Baseline vs. CI)	-0.067	0.19	39	-0.36	0.725	
Time1 × Syntax (Baseline vs. Transitive)	0.237	0.63	51	0.37	0.710	
Time1 × Syntax (Baseline vs. CI)	0.453	0.63	51	0.71	0.479	
Time2 × Syntax (Baseline vs. Transitive)	-1.686	0.77	43	-2.19	0.035	*
Time2 × Syntax (Baseline vs. CI)	0.977	0.77	43	1.27	0.212	
Time3 × Syntax (Baseline vs. Transitive)	0.104	0.58	49	0.18	0.858	
Time3 × Syntax (Baseline vs. CI)	-0.365	0.58	49	-0.63	0.532	

Notes. CI = Conjoined Subject Intransitive; Time1 = linear time; Time2 = quadratic time; Time3 = cubic time. Random effects structure for Subjects included an intercept plus effects for Time1, Time2 and Time3. R-code used with lmerTest: Elogit ~ (Time1 + Time2 + Time3) \* Syntax + (Time1 + Time2 + Time3 | Subject). Degrees of freedom (df) calculated using the Satterthwaite approximation.

**Table 9.** Growth curve analysis: linear mixed effect model of (elogit) two-participant overall looking preferences in Animate-Inanimate contexts including effects of time and syntax (baseline, transitive, CI) for 3- to 4-year-olds.

Fixed effects	Estimate	S.E.	df	t	p-value	Significance
Intercept	-0.063	0.11	43	-0.57	0.571	
Time1	0.061	0.48	49	0.13	0.900	
Time2	-0.386	0.40	55	-0.97	0.334	
Time3	0.552	0.41	61	1.36	0.180	
Syntax (Baseline vs. Transitive)	0.406	0.14	43	2.93	0.005	*
Syntax (Baseline vs. CI)	-0.411	0.14	43	-2.97	0.005	*
Time1 × Syntax (Baseline vs. Transitive)	1.624	0.61	49	2.67	0.010	*
Time1 × Syntax (Baseline vs. CI)	-0.107	0.61	49	-0.18	0.862	
Time2 × Syntax (Baseline vs. Transitive)	-0.379	0.50	55	-0.76	0.451	
Time2 × Syntax (Baseline vs. CI)	0.536	0.50	55	1.07	0.288	
Time3 × Syntax (Baseline vs. Transitive)	-1.129	0.51	61	-2.20	0.032	*
Time3 × Syntax (Baseline vs. CI)	-1.213	0.51	61	-2.36	0.021	*

Notes. CI = Conjoined Subject Intransitive; Time1 = linear time; Time2 = quadratic time; Time3 = cubic time; Random effects structure for Subjects included an intercept plus effects for Time1, Time2 and Time3. R-code used with lmerTest:  $\text{Elogit} \sim (\text{time1} + \text{time2} + \text{time3}) * \text{Syntax} + (\text{time1} + \text{time2} + \text{time3} | \text{Subject})$ . Degrees of freedom (df) calculated using the Satterthwaite approximation.

difference from Baseline being briefer and more variable with time. The growth curve analysis for the Animate-Inanimate condition (Table 9) supported these observations. As can be seen in the table, both syntactic contrasts (Baseline vs. Transitive, and Baseline vs. CI) had a significant effect on the intercept, the former being positive and the latter being negative. Thus, in a model that accounts for temporal dynamics, both syntactic contrasts are reliable in the expected directions. The model also shows the dynamics are different for each sentence type. The syntactic contrast of Baseline vs. Transitive reliably interacted with the linear term and the cubic term whereas the syntactic contrast of Baseline vs. CI reliably interacted only with the cubic term. Thus the model is detecting two inflection points in both experimental conditions relative to the Baseline, with the differences from Baseline arising earlier in the CI condition. Finally, we conducted an additional growth-curve analysis (not shown in the table) that directly compared the two experimental conditions (Transitive vs. CI). This analysis confirmed that the two experimental conditions were indeed different from each other; there was a reliable effect of the syntactic contrast (Transitive vs. CI) on the intercept, and on the linear and quadratic time terms ( $p$ 's < .05).

**Two-year-olds.** For 2-year-old children in the Animate-Animate Condition (Figure 10), the growth curve analysis using a three-level factor of Syntax (Baseline, Transitive, CI) yielded no significant effects or interactions. Thus, the dynamics of the looking preferences in either experimental condition could not be distinguished from the dynamics of the Baseline condition. Interestingly however, a separate analysis comparing the experimental conditions to each other (i.e., the Transitive vs. CI contrast) revealed significant differences consistent with past literature. Specifically, the analysis revealed a marginal effect of the quadratic term on the intercept ( $Estimate = -.58$ ,  $SE = .33$ ,  $t(25) = -1.78$ ,  $p = .09$ ), and a significant effect of the cubic term on the intercept ( $Estimate = .66$ ,  $SE = .30$ ,  $t(25) = 2.19$ ,  $p = .04$ ), indicating an overall nonlinear increase in looks to the two-participant event video through time, which arose independently of Syntax. Moreover, a significant effect of Syntax on the intercept ( $Estimate = -.24$ ,  $SE = .07$ ,  $t(1100) = -3.66$ ,  $p < .01$ ) and a marginally significant effect of Syntax on the linear term ( $Estimate = -.63$ ,  $SE = .32$ ,  $t(1100) = -1.97$ ,  $p = .05$ ) emerged, indicating that two-participant event preference scores were higher for Transitive than CI sentences and that this preference increased with time. Thus consistently with past studies, 2-year-olds are showing some sensitivity, in the expected direction, to the transitive vs. CI syntax contrast. No other main effects or interactions were significant (all  $p$ 's > .10).

Similarly, for 2-year-old children in the Animate-Inanimate Condition (Figure 11), the growth curve analysis using a three-level factor of Syntax (Baseline, Transitive, CI) yielded no significant effects or interactions. However, an analysis of the experimental conditions (Transitive vs. CI) revealed a significant effect of Syntax on the intercept ( $Estimate = -.29$ ,  $SE = .06$ ,  $t(924) = -4.69$ ,  $p < .01$ ), and an effect of Syntax on the linear term ( $Estimate = -.77$ ,  $SE = .30$ ,  $t(924) = -2.57$ ,  $p = .01$ ), again indicating 2-year-olds' sensitivity to the Transitive vs. CI contrast. No other main effects or interactions were significant (all  $ps > .10$ ).

Thus, the time-course analyses from 2-year-olds yield a pattern that is weak but consistent with the past literature. Children in this age range show small but significant sensitivity to the syntactic contrast of Transitive versus CI sentences, but comparisons with parallel preferences in a baseline condition (rarely done in the past literature) suggest that this sensitivity is quite weak and the results should be interpreted with some caution.

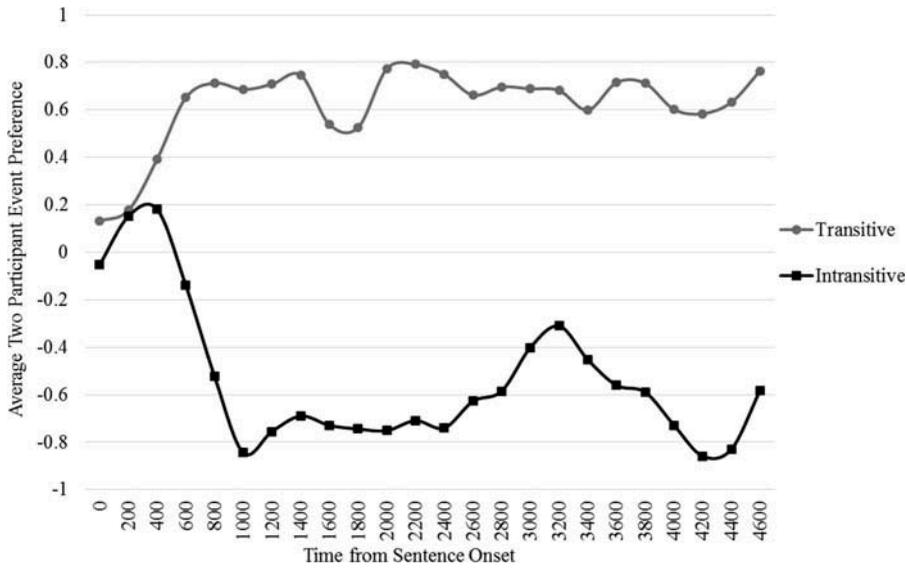
**Adults.** Finally, a parallel growth curve analysis was conducted to analyze adult participants' real time interpretation of transitive and CI sentences.<sup>4</sup> For adults in the Animate-Animate Condition (Figure 12), the growth curve analysis revealed a significant effect of Syntax on the intercept ( $Estimate = -4.60$ ,  $SE = .21$ ,  $t(368) = -22.32$ ,  $p < .01$ ), indicating stronger two-participant event preference scores for Transitive than for CI sentences, a marginal effect of the linear term on the intercept ( $Estimate = -1.06$ ,  $SE = .56$ ,  $t(19) = -1.90$ ,  $p = .07$ ), and a significant effect of Syntax on the linear ( $Estimate = -4.50$ ,  $SE = 1.01$ ,  $t(368) = -4.45$ ,  $p < .01$ ), quadratic ( $Estimate = 5.63$ ,  $SE = 1.01$ ,  $t(368) = 5.57$ ,  $p < .01$ ) and cubic terms ( $Estimate = -4.16$ ,  $SE = 1.01$ ,  $t(368) = -4.12$ ,  $p < .01$ ), indicating that Transitive sentences, compared to CI structures, are associated with a stronger increase through time in two-participant event preference scores.

For adults in the Animate-Inanimate condition (Figure 13), the pattern of results was similar to that observed for the Animate-Animate condition. The growth curve analysis revealed a significant effect of Syntax on the intercept ( $Estimate = -5.11$ ,  $SE = .15$ ,  $t(315) = -34.45$ ,  $p < .01$ ), indicating stronger two-participant event preference scores for Transitive than for CI sentences, a significant effect of the cubic term on the intercept ( $Estimate = -1.05$ ,  $SE = .40$ ,  $t(21) = -2.63$ ,  $p = .02$ ), and a significant effect of Syntax on the linear ( $Estimate = -9.71$ ,  $SE = .73$ ,  $t(315) = -13.36$ ,  $p < .01$ ), quadratic ( $Estimate = 6.35$ ,  $SE = .73$ ,  $t(315) = 8.75$ ,  $p < .01$ ) and cubic terms ( $Estimate = -2.33$ ,  $SE = .73$ ,  $t(315) = -3.20$ ,  $p < .01$ ), indicating that Transitive sentences are associated with a stronger increase through time in two-participant event preference scores than CI sentences.

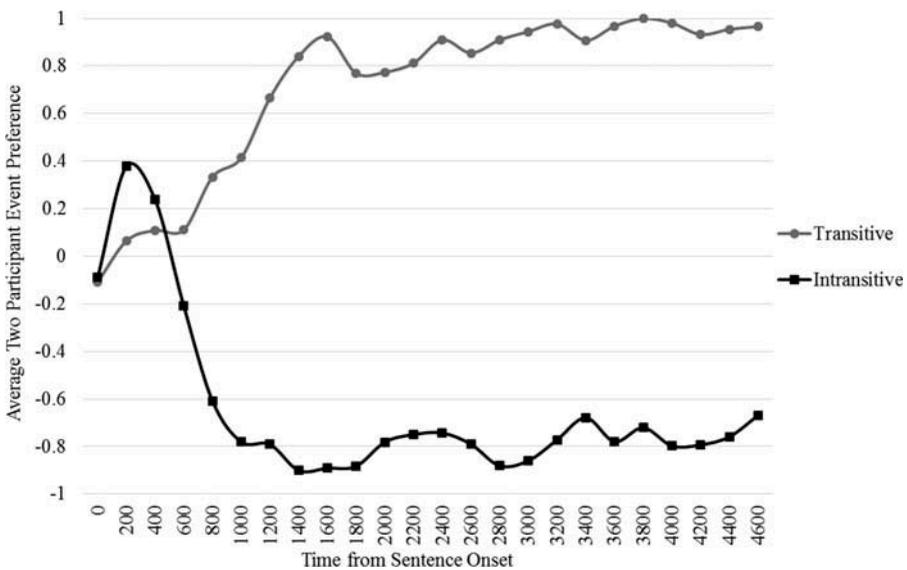
**Summary of time-course analyses.** The results that were obtained from the time-course analyses confirm and extend the results obtained from the overall whole-window analyses: as predicted by the theory of syntactic bootstrapping, all participants, regardless of age and animacy condition, discriminate between transitive and CI sentences, and all but the youngest group of participants show evidence of being able to pair transitive sentences with two-participant events. However, a significant preference for one-participant events when presented with CI sentences emerged and developed through time only for adults and 3- to 4-year-olds in the Animate-Inanimate condition. This result is consistent with the experimental hypothesis that children's reported failure to pair CI sentences with one-participant events is, at least in part, due to the fact that some subevent components of Animate-Animate two-participant events represent good candidates for novel verbs in CI frames.

While the time course analysis provided evidence that 2-year-olds are indeed able to distinguish between Transitive and CI sentences, in line with previous studies in the literature (Arunachalam &

<sup>4</sup>For adults, looks to the two-participant events were already different from chance at the onset of the experimental sentence when the video was presented for the second time; for this reason, we only analyze and show the time plots for the first video presentation (E1).



**Figure 12.** Animate-Animate condition: Adult participants' two-participant overall looking preferences as a function of syntax and time.



**Figure 13.** Animate-Inanimate condition: Adult participants' two-participant overall looking preferences as a function of syntax and time.

Waxman, 2010; Kidd et al., 2001), the present study did not find positive evidence that hearing novel verbs within an informative syntactic frame has a consistent effect on the time that 2-year-olds allocate to the different events, at least when compared with a condition in which children hear neutral syntax (the Baseline condition). Many factors might have contributed to this null effect, among which are the small sample size and high variability in the Baseline condition. Since very few studies in this literature have used a baseline group for comparison (Naigles & Kako, 1993; Noble et al., 2010 being two exceptions), we leave this issue open for further research.

## General discussion

A wealth of work in the tradition of syntactic bootstrapping indicates that young children use the syntactic context in which novel words appear to infer aspects of their meaning. In her classic studies of the effect of syntactic structure on children's verb learning, Naigles (1990) showed that young 2-year-olds consistently map two-argument transitive sentences onto two-participant causal events, and map one-argument intransitive sentences onto one-participant events. The former result has been consistently replicated and extended in subsequent research. However, a much less consistent picture has emerged with respect to children's ability to map plural-subject intransitive sentences onto one-participant events. In very young children, this difficulty might stem from a failure to understand that conjoined NPs are to be parsed as a single argument; then, considering each NP as a separate argument, they use an agent-first strategy to impose an SO analysis on CI sentences (Gertner & Fisher, 2012). In short, their syntactic bootstrapping procedure fails because it is operating on a mistaken parse.

In introductory comments, we suggested another source of difficulty with CI sentences. We hypothesized that even children who are capable of correctly parsing and interpreting both simple transitive sentences ("The boy is glorping the girl") and conjoined subject intransitive (CI) sentences ("The boy and the girl are glorping") might experience some confusion when presented with the latter sentence type under certain experimental conditions. Specifically, in the presence of two side-by-side videos that both involve two actors (a boy and a girl) engaged either in a two-participant causal event (e.g., a boy causing a girl to spin) or two independent one-participant noncausal events (a boy and a girl spinning independently), the sentence "The boy and the girl are glorping" could apply to both videos if 'glorp' meant 'play' or 'smile' for example, rather than 'spin.'

Among 3- to 4-year-old children, we observed looking preferences consistent with this hypothesis. Children in this age range who heard a transitive sentence generated a reliable preference to look at the two-participant causal event over the one-participant noncausal events, as compared to children who heard a neutral sentence ("I see glorping!"). These Baseline subjects showed no preference for either video. In contrast, children who heard a CI sentence generated looking preferences indistinguishable from the Baseline condition, that is, no preference for either depicted video. This suggests that, on average, children in this age range interpreted a sentence like "The boy and the girl are glorping" as matching either video, perhaps because both scenes were compatible with joint activities such as playing or simultaneous independent activities such as smiling. This explanation received further support from the behavior of a separate group of children who instead were asked to choose between two videos involving an animate and an inanimate object (e.g., a boy and an umbrella). Here the corresponding two-participant video depicted a boy causing an umbrella to spin whereas the one-participant video depicted a boy and an umbrella spinning independently. Joint or matched activities of playing and smiling are much less plausible descriptions of the two-participant video (as confirmed by a norming study conducted with adults; see footnote 1), thus potentially blocking a source of semantic ambiguity associated with the CI sentences. Indeed, both sentence types now deviated from the neutral Baseline condition in the expected directions. Transitive sentences (e.g., "The boy is glorping the umbrella") produced a reliable preference for the two-participant video, whereas CI sentences ("The boy and the umbrella are glorping") produced a preference for the one-participant video as compared to Baseline (a difference that approached significance in the overall whole-window analysis and was significant in the growth-curve analyses, which took the temporal dynamics of looking preferences into account).

Two-year-olds provided a more complex and perhaps noisier pattern of looking preferences that was nonetheless consistent with the past literature. Like prior studies, we found some evidence that 2-year-olds are sensitive to the syntactic contrast of transitive vs. CI sentence types. In growth curve analyses of both the Animate-Animate and Animate-Inanimate conditions, transitive sentences generated a reliably greater two-participant preference than CI

sentences. However, this difference was not reliable in the overall whole-window analyses. Moreover, all analyses that compared looking preferences to the Baseline condition yielded no detectable differences; that is, neither transitive nor CI sentences were distinguishable from the Baseline conditions. This suggests some caution in interpreting the transitive versus CI contrasts in this age group, particularly since few past studies have used a baseline condition. Nevertheless the developmental progression indicates that as soon as children show signs of being able to parse and interpret CI sentences, they also show signs of understanding the full range of semantic options consistent with this structure.

A final word needs to be said—though it must be somewhat speculative until further experimentation is done—about why our adult control subjects, unlike the children, quite rigidly obeyed the initially hypothesized mapping between syntax and event structure (e.g., Gleitman, 1990; Naigles, 1990; Fisher et al., 1994); that is, adults preferred two-participant event scenes as the referent for transitive sentences and the one-participant event scenes for CI sentences, despite the fact that, as we have argued, CI sentences with matched animacy plausibly map onto both scenes. Why aren't these mature subjects, as we hypothesized for novices, flummoxed by the myriad potential representations that match one-argument stimulus sentences against the visually perceived world? We suggest that adult experimental participants fall back on strategies that are in some circles deplored as artifacts of “task demands” and in others applauded as triumphs of Gricean cooperativeness (Grice, 1975). When an experimental situation systematically varies a single stimulus opposition, sophisticated participants may discover and seize upon this explicitly manifest opposition as a uniform basis for strategically selecting among their responses. Young children might be less reliant on such inferential conversational principles (e.g., Conti & Camras, 1984; Eskritt, Whalen, & Lee, 2008).

### **Remarks and conclusions**

Concentrating our experimental attentions on one apparently simple sentence type and on some apparently simple structural constraints on its interpretation onto the world, we—following Gertner and Fisher (2012)—have discovered surprisingly complex patterns of participants' responses, depending, *inter alia*, on such variables as the participants' own ages/language knowledge stages, properties of “the world” (caused and uncaused motion and states) that co-occur with speech, and of syntactic type of the interlocutor's speech (NP and argument number). In this sense, the study of CI sentences is a microcosm of the multifaceted task that awaits the novice trying to recover properties of the syntax-semantics interface of the exposure language without yet being anchored securely at either end of this interface.

Specifically, as we have seen here, the recovery of the message being conveyed by one's interlocutor requires knowledge of word meanings and functions (the coordinate interpretation of “and,” the surface manifestations of singular/plural morphology as “is” vs. “are,” and the informal semantics of “boy” and “umbrella”). But even if all this lexical knowledge is in place, the true novice may be controlling only the flat surface deployment of component words and phrases in the sentential string, without sure knowledge of the intended (or language-sanctioned) hierarchical parse: that one lexical item may render a component part of an argument, another of connective properties (bringing into play such possible universal strictures as the “A over A” principle; Chomsky, 1964), and so forth. This is the aspect of the problem that was focused on, manipulated, and elegantly explained in Gertner and Fisher's (2012) study of 2-year-olds confronted with CI sentences.

As we further hypothesized and studied here, the child's puzzle in “solving for English” also includes the fact that the words spoken will map onto the observed world in complex and variable referential and representational ways. Connecting the dots of form and meaning is therefore a task of enormous complexity. It is surely hard to map from a meaning to a structure for these reasons, and hard to map from a structure to a meaning, but the child's problem is particularly hard because he or she may be solving for both of these terms of the equation at once so as to extract the meaning of a single new word.

To illustrate, consider acquiring our favorite word “glorp” while hearing the sentence “Benjamin glorped the cookie on the table.” As the mapping theories (syntactic bootstrapping, Gleitman, 1990; semantic bootstrapping, Pinker, 1984) hypothesize, the observed world partly reins in the intended meaning while the structure itself partly gives considerable further constraint. The problem, though, is that the linear string does not fully determine this structure (any more than “the world” imposes a unique meaning; Chomsky, 1957): if “on the table” is interpreted as an NP-modifier, then the sentence has two arguments and “glorp” might mean “see” or “eat.” But if “on the table” is construed instead as a VP-modifier, then this sentence has three arguments and might mean “put.”<sup>5</sup> So here is the learner’s dilemma: One can radically rein in, thus discover, the meaning of a new word by understanding its syntactic environment, or alternatively if one looks out into the co-occurring world (observing, say, seeing but not eating going on within one’s visual compass), one can impose the structure upon the word-string by matching it to co-occurring events. But what if one cannot solve either side of the equation in a secure way?

In the studies presented here, we accordingly asked about which of these problems—the parsing problem or the world-interpretation problem—was chiefly responsible for young learners’ difficulties in interpreting what are (in the minds of experimenters and other sophisticated English speakers) conjoined nominal one-argument structures representing uncaused events (CIs). Whereas prior investigators provided first traction into this problem by varying the potential syntax (Gertner & Fisher, 2012, who varied NP order to see if it determined, for young subjects, argument order in these sentences; Noble et al., 2010, who varied NP number while keeping notional number unaltered for related reasons), our new studies determined that the existing explanation could not plausibly cover all the facts, for after all why should a linguistically sophisticated 3- or 4-year-old continue to show difficulties that novices might exhibit in the first two years of life? We therefore began to show the influence of the other side of the form-meaning equation by radically reducing the ways that the observed world could be related to the spoken sentence. Except under the most fanciful circumstances, girls and umbrellas do not play together, they absolutely could not smile together, and they do not interact as co-performing partners. Thus “the girl and the boy glorp” is more ambiguous or indeterminate in its interpretation than “the girl and the umbrella glorp.” Language learners respond to this knowledge in the ways one might expect, leaving several semantic options open in the first situation and closed in the second situation.

## Acknowledgments

We would like to thank Annalisa Benetello, Alon Hafri, and Josh “Shwaman” for starring in our videos, and Ann Bungler for recording the experimental sentences. Thanks to Alon Hafri, Yi Lin, Michelle Luong, Elisabeth Trueswell, and Kristina Woodard for their assistance in collecting the eye movement data from children. We also thank the preschools, parents, and children who participated in this study.

## Funding

This work was funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development Grant 1-R01-HD-37507 awarded to Lila R. Gleitman and John C. Trueswell.

## References

- Arunachalam, S., & Waxman, S. R. (2010). Meaning from syntax: Evidence from 2-year-olds. *Cognition*, 114, 442–446.
- Barr, D. J. (2008). Analyzing “visual world” eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59, 457–474.

<sup>5</sup>This problem never fully disappears even for the mature language user, for this is the same problem faced by any greedy real-time interpreter of sentences, especially when confronted with temporary syntactic ambiguity. How much should one rely on the linguistic and the nonlinguistic evidence when committing to a parse when, in the moment, neither is completely determinant? Sophisticated coordination of both sources of evidence appears necessary (e.g., Novick, Thompson-Schill, & Trueswell, 2008; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995).

- Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 173–218). New York, NY: Cambridge University Press.
- Bavin, E. L., & Growcott, C. (2000). Infants of 24–30 months understand verb frames. In M. Perkins & S. Howard (Eds.), *New directions in language development and disorders* (pp. 169–177). New York, NY: Kluwer Academic.
- Bever, T. G. (1970). The cognitive basis for linguistic structure. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–362). New York, NY: John Wiley.
- Bloom, P. (2002). Mindreading, communication, and the learning of the names for things. *Mind and Language*, 17, 37–54.
- Brown, R. (1957). Linguistic determinism and the part of speech. *Journal of Abnormal and Social Psychology*, 55, 1–5.
- Chomsky, C. (1969). *The acquisition of syntax in children from 5 to 10*. Cambridge, MA: The MIT Press.
- Chomsky, N. (1957). *Syntactic structures*. The Hague, the Netherlands: Mouton.
- Chomsky, N. (1964). *Current issues in linguistic theory*. The Hague, The Netherlands: Mouton.
- Chomsky, N. (1993). *Lectures on government and binding: The Pisa lectures* (7th ed.). Berlin, Germany: Mouton de Gruyter. (Original work published 1981)
- Conti, D. J., & Camras, L. A. (1984). Children's understanding of conversational principles. *Journal of Experimental Child Psychology*, 38, 456–463.
- Croft, W. (1991). *Syntactic categories and grammatical relations*. Chicago, IL: University of Chicago Press.
- Dittmar, M., Abbot-Smith, K., Lieven, E. V. M., & Tomasello, M. (2008). German children's comprehension of word order and case marking in causative sentences. *Child Development*, 79, 1152–1167.
- Eskritt, M., Whalen, J., & Lee, K. (2008). Preschoolers can recognize violations of the Gricean maxims. *British Journal of Developmental Psychology*, 26, 435–443.
- Fisher, C. (1996). Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive Psychology*, 31, 41–81.
- Fisher, C. (2002). Structural limits on verb mapping: The role of abstract structure in 2.5-year-olds' interpretations of novel verbs. *Developmental Science*, 5, 56–65.
- Fisher, C., Hall, D. G., Rakowitz, S., & Gleitman, L. (1994). When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua*, 92, 333–375.
- Fisher, C., Gleitman, L. R., & Gleitman, H. (1991). On the semantic content of subcategorization frames. *Cognitive Psychology*, 23, 331–392.
- Gertner, Y., & Fisher, C. (2012). Predicted errors in children's early sentence comprehension. *Cognition*, 124, 85–94.
- Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: abstract knowledge of word order in early sentence comprehension. *Psychological Science*, 17, 684–691.
- 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, 23–64.
- Göksun, T., Küntay, A., & Naigles, L. (2008). Turkish children use morphosyntax in extending verb meaning. *Journal of Child Language*, 35, 291–323.
- Grice, P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics* 3 (pp. 41–58). New York, NY: Academic Press.
- Hirsh-Pasek, K., Golinkoff, R. M., & Naigles, L. (1996). Young children's use of syntactic frames to derive meaning. In K. Hirsch-Pasek & R. Golinkoff (Eds.), *The origins of grammar: Evidence from early language comprehension* (pp. 123–158). Cambridge, MA: The MIT Press.
- Kidd, E., Bavin, E. L., & Rhodes, B. (2001). Two-year-olds' knowledge of verbs and argument structures. In M. Almgren, A. Barreña, M. J. Ezeizabarrena, I. Idiazabal, & B. MacWhinney (Eds.), *Research on child language acquisition: Proceedings of the 8th conference of the international association for the study of child language* (pp. 1368–1382). Somerville, MA: Cascadilla Press.
- Landau, B., & Gleitman, L. R. (1985). *Language and experience: Evidence from the blind child*. Cambridge, MA: Harvard University Press.
- Langacker, R. (1987). *Foundations of cognitive grammar, 1*. Stanford, CA: Stanford University Press.
- Lee, J., & Naigles, L. (2008). Mandarin learners use syntactic bootstrapping in verb acquisition. *Cognition*, 106, 1028–1037.
- Lidz, J., Gleitman, H., & Gleitman, L. (2003). Understanding how input matters: Verb learning and the footprint of universal grammar. *Cognition*, 87, 151–178.
- Lukyanenko, C., & Fisher, C. (2012). *Agreeing with conjoined subjects: 3-year-olds know one and one are (usually) two*. Paper presented at the Annual Boston University Conference on Language Development, Boston, MA.
- Mirman, D. (2014). *Growth curve analysis and visualization using R*. Boca Raton, FL: Chapman and Hall/CRC.
- Mirman, D., Dixon, J. A., & Manguson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59, 475–494.
- Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, 17, 357–374.
- Naigles, L. (1996). The use of multiple frames in verb learning via syntactic bootstrapping. *Cognition*, 58, 221–251.

- Naigles, L., & Hoff-Ginsberg, E. (1995). Input to verb learning: Evidence for the plausibility of syntactic bootstrapping. *Developmental Psychology*, 31, 827–837.
- Naigles, L., Gleitman, G., & Gleitman, L. (1993). Children acquire word meaning components from syntactic evidence. In E. Dromi (Ed.), *Language and cognition: A developmental perspective*, 5 (pp. 104–140). Norwood, NJ: Ablex Publishing.
- Naigles, L. G., & Kako, E. T. (1993). First contact in verb acquisition: Defining a role for syntax first contact in verb acquisition. *Child Development*, 64, 1665–1687.
- Nappa, R., Wessel, A., McEldoon, K. L., Gleitman, L. R., & Trueswell, J. C. (2009). Use of speaker's gaze and syntax in verb learning. *Language Learning and Development*, 5, 203–234.
- Noble, C. H., Rowland, C. F., & Pine, J. M. (2011). Comprehension of argument structure and semantic roles: Evidence from English-learning children and the forced-choice pointing paradigm. *Cognitive Science*, 35, 963–982.
- Noble, C., Theakston, A. L., & Lieven, E. V. M. (2010). *Comprehension of intransitive argument structure: The first noun as causal agent bias*. Paper presented at the Annual Boston University Conference on Language Development, Boston, MA.
- Novick, J. M., Thompson-Schill, S. L., & Trueswell, J. C. (2008). Putting lexical constraints in context into the visual world paradigm. *Cognition*, 107, 850–903.
- Papafraou, A., Cassidy, K., & Gleitman, L. R. (2007). When we think about thinking: The acquisition of belief verbs. *Cognition*, 105, 125–165.
- Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
- Pulverman, R., Capote, K., Hughson, S., Garber, W., & Sorrell, P. (2012). *Syntactic bootstrapping does not equal word learning: What do toddlers think a novel verb means?* Poster presented at the Annual Boston University Conference on Language Development, Boston, MA.
- Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.
- Sheline, L., Waxman, S. R., & Arunachalam, S. (2013). *Understanding conjoined-subject intransitives: Two-year-olds perform like adults*. Poster presented at Society for Research in Child Development, Seattle, WA.
- Slobin, D. I., & Bever, T. G. (1982). Children use canonical sentence schemas: A crosslinguistic study of word order and inflections. *Cognition*, 12, 229–265.
- Tager-Flusberg, H., de Villiers, J., & Hakuta, K. (1982). The development of sentence coordination. In S. Kuczaj (Ed.), *Language development I* (pp. 201–243). Hillsdale, NJ: Erlbaum.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.
- Trueswell, J. C., Kaufman, D., Hafri, A., & Lidz, J. (2012). Development of parsing abilities interacts with grammar learning: Evidence from Tagalog and Kannada. In A. K. Biller, E. Y. Chung, & A. E. Kimball (Eds.), *Proceedings of the 36th Annual Boston University Conference on Language Development* (pp. 620–632). Somerville, MA: Cascadilla Press.
- Yuan, S., & Fisher, C. (2009). “Really? She blinked the baby?”: Two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science*, 20, 619–626.