Language Learning and Development

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/hlld20

Thematic Content, Not Number Matching, Drives Syntactic Bootstrapping

Laurel Perkins, Tyler Knowlton, Alexander Williams & Jeffrey Lidz

To cite this article: Laurel Perkins, Tyler Knowlton, Alexander Williams & Jeffrey Lidz (19 Jul 2024): Thematic Content, Not Number Matching, Drives Syntactic Bootstrapping, Language Learning and Development, DOI: 10.1080/15475441.2024.2372663

To link to this article: https://doi.org/10.1080/15475441.2024.2372663

Published online: 19 Jul 2024.

LANGUAGE LEARNING AND DEVELOPMENT
https://doi.org/10.1080/15475441.2024.2372663

Routledge
Taylor & Francis Group

Thematic Content, Not Number Matching, Drives Syntactic Bootstrapping

Laurel Perkins[a], Tyler Knowlton[b], Alexander Williams[c,d], and Jeffrey Lidz[c]

[a]Department of Linguistics, University of California, Los Angeles, USA; [b]MindCORE, University of Pennsylvania, Pennsylvania, USA; [c]Department of Linguistics, University of Maryland, College Park, USA; [d]Department of Philosophy, University of Maryland, College Park, USA

ABSTRACT

Children use correlations between the syntax of a clause and the meaning of its predicate to draw inferences about word meanings. On one proposal, these inferences are underwritten by a structural similarity between syntactic and semantic representations: learners expect that the number of clause arguments exactly matches the number of participant roles in the event concept under which its referent is viewed. We argue against this proposal, and in favor of a theory rooted in syntactic and semantic contents – in mappings from syntactic positions to thematic relations. We (i) provide evidence that infants view certain scenes under a concept with three participant relations (*a girl* taking *a truck* from *a boy*), and (ii) show that toddlers do not expect these representations to align numerically with clauses used to describe those scenes: they readily accept two-argument descriptions ("she pimmed the truck!"). This argues against syntactic bootstrapping theories underwritten by mappings between structural features of syntactic and semantic representations. Instead, our findings support bootstrapping based on grammatical and thematic content. Children's earliest inferences may rely on the assumption that the syntactic asymmetry between subject and object correlates with a difference in how their referents relate to the event described by the sentence.

Introduction

There are correlations between the syntax of a clause and the meaning of its predicate (e.g. Dowty, 1991; Fillmore, 1968, 1970; Jackendoff, 1972; Levin & Hovav, 2005). These correlations might be used to infer one from the other. If a learner infers something about a word's meaning from its syntax, we call that inference syntactic bootstrapping (Gleitman, 1990; Landau & Gleitman, 1985). One prominent proposal in the syntactic bootstrapping literature holds that, at least at the earliest steps into language learning, these inferences are underwritten by a wholly structural similarity between syntactic and semantic representations. This view holds that the number of "argument places" in a clause exactly matches the number of "participant roles" in the event concept under which its referent event is viewed (Fisher, 1996; Lidz & Gleitman, 2004; Naigles, 1990; Yuan et al., 2012; inter alia). On this view, what matters at both levels is a shallow structural property, the number of slots in a representation. In this paper, we raise doubts about this variant of the syntactic bootstrapping hypothesis, and argue instead for a theory rooted in correspondences between, not structures, but their contents: specifically, in mappings from syntactic positions to "thematic relations."

The content-based bootstrapping theory that we argue in favor of here is partially motivated by earlier doubts about the empirical support for the structure-matching account. In particular,

CONTACT Laurel Perkins perkinsl@ucla.edu Department of Linguistics, University of California, 3125 Campbell Hall, Los Angeles, CA 90095, USA

earlier studies find toddlers accepting intransitive clauses, with one syntactic argument place, as descriptions of scenes that experimenters expected to be viewed as two-participant events (Arunachalam & Waxman, 2010; Noble et al., 2011; Yuan et al., 2012). These findings seem at odds with the predictions of one-to-one matching between arguments and participants. But prior work swept these doubts away by retracting the expectation, plausible but not experimentally justified, that toddlers would view the given scene as a two-participant event. If they did not, then it is no argument against the structure mapping proposal that they allowed the scene to be described with a one-argument clause. We address this concern, and find that the original doubts are nonetheless potent.

We first take steps to diagnose the structure of the event representations under which children view our stimulus scenes, and then show that they do not require these representations to align numerically with clauses used to describe those scenes. This finding argues against bootstrapping theories underwritten by mappings between structural features of syntactic and semantic representations, but is consistent with mappings between syntactic positions and semantic relations: children's earliest bootstrapping inferences may rely on the assumption that the syntactic asymmetry between subject and object correlates with a difference of semantic content, in how their referents relate to the event described by the sentence. The conclusion is welcome, we believe, since it allows the learning heuristics of the child to be immediately consistent with the facts of adult languages, where the surface arguments in a clause may be fewer in number than the participant roles in the concept it expresses.

Two types of correspondence theories

Previous bootstrapping literature has debated which forms of evidence, syntactic or conceptual, can most reliably be used a basis of inference for children with immature knowledge of grammar and word meanings (Fisher, 1996; Fisher et al., 2019; Gleitman, 1990; Pinker, 1984, 1989, 1994, inter alia). Our focus in this paper is a different question, orthogonal to this debate. Here, we aim to diagnose the specific correspondence relations that underwrite early bootstrapping inferences – knowledge that is independent of the direction in which these inferences are performed (from syntax to semantics, or vice versa). We make the assumption, common to both bootstrapping theories, that these inferences relate structured syntactic and conceptual representations. Our question is how children expect these representations to correspond to each other when mapping between sentences they hear and scenes they perceive.

On one family of hypotheses, which we will call *Thematic Linking*, children expect correspondences between the specific contents of their syntactic and conceptual representations. They might expect, for instance, that in transitive clauses, subjects tend to name agents while objects tend to name patients, if the clause describes an action (Dowty, 1991; Pinker, 1984, 1989; Williams, 2015). This is similar in kind to an expectation that toddlers acquiring English seem to have about nouns versus verbs (He, 2015; He & Lidz, 2017; Waxman, 1999; Waxman & Booth, 2001): verbs tend to classify events, while nouns tend to classify objects. In both cases a distinction in the content of a syntactic representation (subject/object, noun/verb) is linked to one in the content of a conceptual representation (agent/patient, object/event).

On another influential hypothesis, which we will call *One-to-One Matching*, the correlation is instead wholly structural. Children expect that the syntactic arguments they perceive in a clause should align one-to-one with the participant roles of a concept under which they view the event it describes; that is, the two should match in number (Fisher et al., 2019; Lidz & Gleitman, 2004; Naigles, 1990). These initial expectations may or may not resemble the knowledge that mature speakers have of their language. These hypotheses thus have different implications for the developmental steps that learners take on the way to the adult grammar.

In this section, we examine the motivations and challenges for each of these hypotheses, and show how both can account for the existing behavioral data on early verb learning. This motivates the need for a novel test to empirically differentiate the two hypotheses.

Arguments and participants

Before that, we need to say how we will use important terms, especially "argument" and "participant." By "argument" we will mean something wholly syntactic. An "argument" is a phrase in one of the clause-level syntactic relations that grammarians call subject, direct object and indirect object relations, collectively the "argument relations." We commit to no further analysis of these notions. The terms "subject" and "object" will serve as useful descriptive rubrics for broad generalizations, whether or not they label categorical primitives in the syntactic representations of either children or adults. In particular, it may suffice that children perceive a syntactic asymmetry between what we call subjects and objects, whether or not all subjects have certain properties that distinguish them intrinsically.

We use "participant" for an individual represented as being in a "participant relation" to an event. A participant relation is among those entailed by an event concept: those an event must enter if it falls under the concept, just as an object that falls under CYLINDER must be related to its height and diameter. But not every entailed relation can be a participant relation. Otherwise it would be strange to even consider whether arguments might match participants in number. This just could not be true, since the number of arguments tops out around four, while event concepts generally entail a great many relations. Take an event that falls under SINGING. It will not only have a singer *a* and something *b* we can say was sung, but will also happen at a time *c* and a location *d*, with some part *e* of the singer *a* moving in a certain way *f* within a fluid *g* for duration *h* so as to cause a pattern *i* of compression within *g* that materially instantiates what was sung, and so on. Thus the participant relations must be distinguished somehow from among all those that are entailed. We take the distinction to be psychological: the participant relations are those an event concept represents explicitly, or makes salient, when it is tokened (Wellwood et al., 2015; Williams, 2015).

For convenience we formalize this in terms of the valence (adicity, arity, degree) of the concept. A one-place event concept, true or false of an event *e*, has no participant roles; a two-place concept, true or false of an event and an object (*e*,*x*), has one participant role; and so on. To show in our notation that a concept has valence *n* we give it a name with *n* parts, separated by hyphens. Each part is an informal label for the role associated with one term in the relation. Thus SINGING names a one-place concept, with no participant roles, whose only term is a singing, while SINGING-AGENT names a two-place concept with one participant role, informally labeled AGENT, instantiated by the singer.[1] Importantly, concepts whose name shares the same first part, such as SINGING and SINGING-AGENT, share all the same entailments for their event, in just the same sense that "*x* is a cylinder" shares the entailments of "*x* is a cylinder with a height and a diameter." They differ only in which entailed relations they link to "slots" in their valence, distinguishing them as participant relations, according to our convention.

One last distinction should be made. If a *clause* expresses an *n*-place concept, it doesn't follow this same concept is also the meaning of the *verb*. The concept may derive from the parts of the clause in various ways, to be discovered by semantic analysis. Thus a theorist who relies on clausal arguments in making inferences about sentence meaning needn't say anything about the valence of the concept expressed by the verb.

Thematic linking

Thematic Linking hypotheses propose that children exploit cross-linguistically reliable correspondences between grammatical and thematic relations (e.g., Pinker, 1984, 1989). Here, we abstract away from the details of how these correspondences should be characterized (Baker, 1988, 1997; Carter, 1988; Dowty, 1991; Fillmore, 1968, 1970; Foley & Van Valin, 1984; Jackendoff, 1972; Levin & Hovav,

[1]We do not require further that parts of the name correspond to parts of the named concept, only that their number matches its valence.

2005; Pearl & Sprouse, 2019; Wechsler, 1995). Some of the most robust patterns across languages might be summarized coarsely as follows (Williams, 2015):

(1) a. In basic transitive clauses describing an action, subjects tend to name agents, and objects tend to name patients.
b. Clauses describing a change tend to realize the thing being changed.
c. Clauses describing an action tend to realize the agent of the action.

Generalization (1a) captures the tendency of transitive action verbs to occur in basic clauses where the subject names the agent of the action and the object names the patient, rather than with the reverse linking pattern. For instance, verbs that mean "kick" do not tend to occur in sentences like **The table kicked Kim*, in which *the table* names the thing kicked and *Kim* names the kicker. Generalization (1b) captures cross-linguistic tendencies for verbs describing events of change to distribute with arguments expressing the patient of that change. For example, the English verb *break* can occur in a clause whose sole argument names the thing that got broken (*The vase broke*), but not in a clause whose sole argument names the breaker (**Kim broke*, with the intended meaning "Kim broke something"). Generalization (1c) captures the tendency of verbs describing an action of an agent to distribute with arguments expressing that agent, though this tendency is much weaker than the first two. For example, the English verb *sweep* can occur in a clause whose sole argument names the sweeper (*Kim swept*), but not in a clause whose sole argument names the thing that got swept (**The floor swept*, with the intended meaning "The floor got swept").

One argument for Thematic Linking as a bootstrapping hypothesis is its ability to account for these cross-linguistic generalizations (Baker, 1997; Dowty, 1991; Pinker, 1984, 1989). Suppose that young learners have knowledge of the sorts of principles in natural language grammars that give rise to the generalizations in (1). Those initial expectations would tend to be consistent with the specific argument structure patterns in the language that they are acquiring. Moreover, those expectations could help explain why similar patterns are exhibited in diverse languages: they arise from grammatical principles that form part of learners' initial linguistic capacities, and thus shape the grammatical inferences that learners draw from their data.

Prior empirical support

Prior tests of young children's novel verb learning provide suggestive support for Thematic Linking. Toddlers by 17–24 months show sensitivity to the canonical word order of their language and awareness of how subjects and objects link to different participant roles (Gertner et al., 2006; Hirsh-Pasek & Golinkoff, 1996; Lidz et al., 2017). For instance, Gertner et al. (2006) found that 21- to 24-month-olds who heard *The duck is gorping the bunny* preferred a scene in which a duck pushed a bunny, over a scene in which the bunny pulled the duck. These toddlers appeared to link the argument in subject position to the agent in an event, and the argument in object position to the patient.

Slightly older toddlers have been found to draw inferences from the thematic content of intransitive subjects (Bunger & Lidz, 2004, 2008; Scott & Fisher, 2009), which correlates with differences in event type, as characterized above in (1b-c). In Scott and Fisher (2009), 28-month-olds inferred that an intransitive clause with an inanimate subject (e.g. *The pillow dacked*) was more likely to describe an event in which a change was effected on a patient. Conversely, they inferred that an intransitive clause with an animate subject (e.g. *He dacked*) was more likely to describe an action of an agent that does not effect a change. These toddlers appeared to use animacy, a loose correlate of thematic relations, to guess whether the subject is a likely agent or patient, licensing different inferences about the type of event that the sentence describes.

Useful for initial bootstrapping?

These prior results do not tell us whether Thematic Linking guides verb learning at its very earliest stages. Its viability for children's initial bootstrapping rests on two assumptions about their conceptual and linguistic representations. First, learners at very early stages of development need to perceive the world under event concepts that support the requisite distinction among participant relations, such as agents versus patients. Second, they need some way to reliably identify the requisite distinction among grammatical relations, such as subjects versus objects, in sentences of their language. Note that just as we use "subject" and "object" as useful descriptive rubrics for the relevant asymmetry in grammatical relations, "agent" and "patient" here serve only as descriptive rubrics for the relevant difference in participant relations, whether or not these terms label primitives in children's or adults' conceptual representations. So when we say that learners must represent event concepts that support a distinction between agents and patients, what we require is only that learners represent a relevant asymmetry between participant relations, under which certain participants are viewed as more or less agentive than others (Dowty, 1991).

This first assumption is likely borne out (for a review, see Rissman & Majid, 2019). Adults are able to perceive asymmetries among event participants on the basis of extremely brief visual exposure, indicating that participant relations may be encoded at very early stages of human visual processing (Hafri et al., 2013; Wilson et al., 2011). Infants as young as 6–12 months likewise represent asymmetries among agents and patients (Csibra et al., 2003; Kuhlmeier et al., 2003; Leslie & Keeble, 1987; Muentener & Carey, 2010; Saxe et al., 2005; Woodward, 1998; Yin & Csibra, 2015). Thus, it is likely that young learners' nonlinguistic conceptual representations are rich enough to support bootstrapping to and from specific participant relations.

There is less evidence, however, bearing on the richness of infants' syntactic representations at the earliest stages of verb learning. The generalizations in (1) are cross-linguistically robust because they are stated in terms of grammatical rubrics, like "subject" and "object," which are not directly observable: languages mark the syntactic relations that underlie these rubrics in highly variable ways. If these generalizations are to be explained by the same linking principles used to bootstrap verb meanings, learners would need evidence for subjects and objects in perceived features of the sentences that they hear. This strategy does not require that these relations be represented at the same level of richness as in the adult grammar, or be recognized accurately in all cases: a learner may still be able to employ these linking principles using rougher proxies for grammatical relations, or oppositions among them, such as linear word order. For example, an English learner might infer that a pre-verbal noun phrase is a likely subject and a post-verbal noun phrase is a likely object. But this means that learners would still need to know the language-specific cues that mark *likely* subjects or *likely* objects.

Pinker (1984, 1989, 1994) and Fisher (1996) argue that this knowledge arises only after children acquire some transitive verb meanings. Learners might then use the linking principles in the opposite direction, from semantics to syntax, to infer that the argument naming the agent of the event of the verb is likely the subject, and the argument naming the patient is the object. However, other mechanisms for identifying these core clause arguments may be available. For instance, children might be able to use prosodic cues to syntactic structure, together with frequently-occurring function words, to build an initial syntactic skeleton in which subjects are differentiated from predicates and other arguments (Christophe et al., 2008; de Carvalho et al., 2019; Morgan, 1986; Morgan & Demuth, 1996). Other information, such as asymmetries in noun-verb orders or agreement marking, may serve as the basis for distributionally-based inferences about the canonical position of subjects vs. objects (Maitra & Perkins, 2023; Perkins & Hunter, 2023). Empirically, it remains an open question when and how young infants are able to represent a syntactic asymmetry between the core arguments in a clause. This ability will determine how early a strategy like Thematic Linking could be used.

One-to-one matching

If infants at the onset of verb learning cannot yet differentiate the grammatical relations of clause arguments, then they would need a way to relate coarser properties of their linguistic representations to properties of their event representations. A prominent alternative hypothesis proposes that the relevant property is the number of clause arguments that they represent. If children guess that noun phrases are likely candidates for arguments, then this minimal information might be used to draw an inference about how a clause represents its event, if they moreover expect that arguments will match one-to-one with the participant roles in that event representation (Fisher et al., 2019; Lidz & Gleitman, 2004; Naigles, 1990; Yuan et al., 2012). Put one way, learners might expect that "every participant in an event as it is mentally represented shows up as a syntactic phrase in a sentence describing that event" (Lidz & Gleitman, 2004). Assuming that the implied number of participant roles distinguishes some kinds of events from others – that is, if only some kinds of events can readily be viewed as having that number of participant roles – this could give infants a rough way to narrow down the kinds of events that the clause describes, and thereby to narrow down the possible meanings of a new verb:

> . . . syntactic bootstrapping begins with an unlearned bias toward one-to-one mapping between nouns in sentences and participant roles in conceptual representations. Given this bias, the number of nouns in a sentence is inherently meaningful: Even a young child can infer that a verb combined with two nouns implies two participant roles, whereas a verb combined with one noun implies one participant role. (Yuan et al., 2012)

This proposal goes by various names, including "Structure-Mapping" (Fisher et al., 2019) and "Participant-Argument Matching" (He, 2015; Williams, 2015). For purposes of the current discussion, we call it "One-to-One Matching" (Brandone et al., 2006): it proposes that children's initial bootstrapping inferences are based on one-to-one alignment between the number of variables in the structures under which they represent sentences and scenes, rather than more specific information about the syntactic position of arguments or the content of the participants' relation to their event.

Utility for initial bootstrapping

Because One-to-One Matching only requires the ability to identify noun phrases-as-clause-arguments, but not to identify the syntactic or thematic relations of those arguments, it has the potential to be a powerful bootstrapping strategy even for learners with immature knowledge of their language. But its utility to guide early bootstrapping rests on several assumptions. First, children must be able to reliably perceive the number of arguments in a clause. One-to-One Matching will provide incorrect advice if young learners mis-identify noun phrases as core arguments when they are not (Gertner & Fisher, 2012), or fail to recognize arguments that are present in non-canonical word orders (Gagliardi et al., 2016; Perkins & Lidz, 2020, 2021). Thus, these types of mismatches would need to be relatively rare in a young child's experience, or children would need ways to avoid being misled by them (Perkins et al., 2022).

Moreover, One-to-One Matching requires children to represent scenes in the world under event concepts that are similar to adults' in their participant structure, or they will be unable to use this strategy to identify which events are described by their caregivers' sentences. And importantly, children viewing a particular event must not be likely to shift between conceptual representations that have different numbers of participants but are otherwise equivalent in their content, in exactly the sense that "is a cylinder" and "is a cylinder with a diameter" are equivalent in content (Wellwood et al., 2015; Williams, 2015). Suppose a child can flexibly represent a particular "pushing" scene as either a pushing with two participants, a relation between a pusher and the thing pushed (PUSHING-PUSHER-PUSHED, "a pushing by x of y"), or a pushing with only one participant, a predicate true of anything pushed (PUSHING-PUSHED, "a pushing of y"). We mean these two representations to be mutually entailing, true in all of the same circumstances: they would both be made true by pushings, and differ only in whether something that is necessarily involved in pushings (the pusher) is represented explicitly. If a child readily "down-shifts" in this manner, then One-to-One Matching would be much less useful for deciding which event out of many a particular sentence describes. A transitive

clause might describe the 2-participant representation of this event, or any other event that can be readily down-shifted to a 2-participant representation. An intransitive clause might describe the 1-participant representation of this event, or any other event that can be readily down-shifted to a 1-participant representation. In order for One-to-One Matching to provide useful guidance about the mapping between verb meanings and events, children's event representations must be stable and similar to adults,' and children's expectations must be formulated in terms of the number of participants in events as we readily perceive them (Wellwood et al., 2015; Williams, 2015).

Prior empirical support and challenges

One-to-One Matching receives support from robust evidence that infants as young as 15 to 19 months prefer an event intended to be perceived with two participants when they hear a transitive clause (Arunachalam & Waxman, 2010; Arunachalam et al., 2013; Jin et al., 2014; Messenger et al., 2015; Yuan & Fisher, 20092014; Yuan et al., 2012). These results build on a seminal finding by Naigles (1990), where toddlers who heard *The duck is gorping the bunny* looked longer at a scene in which a duck pushed a bunny, compared to a scene in which the duck and bunny each wheeled their arms. Infants who heard an intransitive sentence (*The duck and the bunny are gorping*) showed the reverse preference. One-to-One Matching accounts for these results in the following way. Children who represent *The duck is gorping the bunny* as having two noun phrase arguments should expect this clause to describe an event that they perceive as having two participants. They should then preferentially map this sentence to the pushing but not the arm-wheeling scene, if only the former is perceived under a two-participant concept.

A complication arises from children's behavior with intransitive clauses. In studies beyond Naigles (1990), infants who hear novel verbs in intransitive frames do not show a reliably above-chance preference for events intended to be viewed with one participant as opposed to two (Arunachalam & Waxman, 2010; Noble et al., 2011; Yuan et al., 2012). Although these results are not predicted under One-to-One Matching, the hypothesis can be salvaged by appealing to indeterminacy in children's event representations. It is possible that infants did not reliably perceive the presented scenes with the number of participants intended by the experimenters. A scene intended to be viewed as one person pushing another might also be viewed under different 1-participant concepts, e.g. one person bending or two people playing (Arunachalam et al., 2016; Brandone et al., 2006; Pozzan et al., 2015). If so, then One-to-One Matching would tell infants that this scene could be described by an intransitive clause.

But this behavior also reinforces a more serious concern about One-to-One Matching as a useful bootstrapping strategy: the expectation of one-to-one correspondence between participants and arguments, for any number, may simply be too strong. Indeed, adopting this assumption would seem to warrant incorrect inferences with some regularity, since there sometimes seem to be fewer arguments than participants. A clause that seems intransitive, at least to a learner unaware of what its verb means, is able to express what surely is not a 1-participant concept. Short passives, like *The truck was moved*, are a familiar first example. These have only one argument on the surface. And to learners who do not yet know the meaning of the verb, there can be no evidence for an inaudible second, until they come to recognize the significance of the passive morphology.[2] Until that point, then, One-to-One Mapping will advise the learner that the verb in the passive clause expresses a 1-participant event concept, even though the same verb in the active would express a 2-participant concept.

The challenge does not end once knowledge of the passive is acquired. Even in basic active clauses of English, with no morphological sign of a "missing" or "demoted" argument, it is not obvious that every perceived participant is always realized as an argument, much less one that can be detected without semantic evidence. A scene in which a girl takes a truck from a boy might plausibly be viewed under a 3-participant concept, one in which the girl, truck, and boy fill participant roles. But it also

[2]The entailed "agent" role of a passive does not have the discourse pragmatics of a silent subject pronoun in languages like Spanish or Mandarin, where these refer to salient or topical referents in the conversation. For passives, therefore, evidence for a silent second NP cannot come from the pragmatics of topic-continuation either, as it presumably does in cases of "pro-drop."

seems easy to describe this scene verbally without naming the boy: *The girl took the truck*. Does this imply that the speaker sees the event under a concept in which the boy is not explicitly represented as a participant (Tatone et al., 2015)? The literature has often assumed this to be the case. But plausibly, the speaker may also have intended this 2-argument clause as a description of the event seen as a 3-participant TAKING, and has simply chosen not to mention one of the participants that are explicit in the event representation: we exclude the boy from the conversation, but not his role in the event from how we conceptualize it. If that is right, then One-to-One Matching provides bad advice in this context. A learner who does not yet know the meaning of *take* would expect that *The girl took the truck* cannot express a 3-participant TAKING, but instead expresses an event viewed as having only 2 participants, different from what the speaker had intended.

This sort of problem becomes even more salient once we broaden our survey of languages beyond English. Many languages allow predicates which entail an agent – and which in English would require transitive syntax, with meanings like "crush" or "repair" – to occur in simple intransitive clauses without an argument noun phrase realizing the entailed agent, and without anything like passive marking. Examples like this can be found in Mandarin, Igbo, Fijian and Hindi (Williams, 2015). Extreme cases include Musqueam and St'át'imcets, in which nearly every verb root can occur, in its bare form, in a simple intransitive clause (Davis, 2010; Davis & Demirdache, 2000; Suttles, 2004). Consider the following example from St'át'imcets:

(2) Qáḿt k^{w}sk^{w}ímčxen
hit.with.projectile det.NAME
Approximately: "K^{w}ímčxen got beaned" (Davis, 2010).

This entails that someone got hit with a thrown object, but the clause has neither the hitter nor the thrown object as syntactic arguments. It is intransitive, and realizes only the role of the person hit (Davis, 2010; Davis & Demirdache, 2000). The other roles are not tied to silent pronominal arguments: to be felicitous, (2) does not require that a thrower or a projectile be topical in the conversation. Nor are they demoted through passivization: the language does have a passive, but it is marked by verbal affixes absent from (2). And yet if a particular "beaning" event is perceived with a hitter, person hit, and thrown object as participants, then One-to-One Matching would erroneously tell learners that a sentence like (2) cannot describe this event. This would lead to incorrect inferences about the meaning of the verb in this sentence. Because intransitive contexts are available to nearly every verb in the language, the assumption of one-to-one correspondence would provide unhelpful advice for St'át'imcets learners in general (Williams, 2015).

Thus, One-to-One Matching not only faces empirical challenges in prior tests of children's verb learning, but it also seems not to characterize the relation between the participant structure of concepts and the argument structure of clauses in adult grammars, and certainly does not characterize the perceived relation for learners unaware of the meaning of the clause. There are many possible responses to these concerns. Perhaps the strategy only serves as an initial verb-learning heuristic, which is abandoned as children acquire more sophisticated knowledge of their language (Fisher, 1996; Lidz & Gleitman, 2004; Lidz et al., 2003). This means that the pathway to a mature grammar involves an errorful stage in which children's early beliefs about the correspondence between syntax and semantics do not resemble those of adults. Hence, this requires a further account of how children's grammatical knowledge develops past this stage.

Another possible response is to weaken the structure-mapping hypothesis by relaxing the one-to-one correspondence assumption. Consider instead an account that we might call "Unidirectional Mapping": perhaps children expect that clause arguments each name a participant, but each event participant does not need to be realized as a clause argument (Fisher et al., 2019). That is, children in general expect that noun phrases are arguments, and furthermore that arguments refer to particular relations that are part of the events described by the verb. This is an important and necessary

assumption under any form of syntactic bootstrapping. However, if young learners do not have any other information about the specific participant relations that particular clause arguments might be naming, then this assumption provides a far weaker basis for inferring verb meaning solely on the basis of argument number, compared to the One-to-One strategy. A learner who hears a transitive clause could infer that it does not describe an event seen with only one participant. But an intransitive clause could describe an event seen with one participant, or two, or three: argument number provides very little guidance on its own, and inferences about the meaning of the verb in this clause would need to be drawn from another source of information, such as animacy, that would allow learners to infer the thematic relation of the clause argument (Williams, 2015). Hence, relaxing the One-to-One assumption implies that early bootstrapping does not rely solely on the structural features of syntactic and conceptual representations. Under the Unidirectional Mapping hypothesis, learners would need access to a further content-based strategy – under the umbrella of hypotheses that we call Thematic Linking – whereby richer sources of syntactic or non-syntactic information, in addition to argument number, help limit the range of event concepts that the clausal predicate expresses (Fisher et al., 2019).

We posit that Thematic Linking on its own provides a simple and parsimonious account for both children's early sensitivities to thematic content and their behavior with transitive vs. intransitive clauses. If infants are able to approximately identify which argument in a transitive clause is the subject and which is the object, and moreover expect subjects to name agents and objects to name patients, then they will prefer an event readily viewed as an agent acting on a patient, over an event readily viewed as only having an agent. And when infants are indifferent between these two event types when hearing an intransitive clause, this may be because they do not know whether the intransitive subject is intended to label an agent of an action vs. a patient undergoing a change. If infants' earliest bootstrapping inferences are content-based in this way, rather than wholly structure-based, we argue that this has a welcome theoretical consequence: it allows a theory of bootstrapping that from the outset is guided by expectations about syntax-semantics correspondence relations consistent with those in the mature grammar of any language.

This paper

Although prior work raises concerns about One-to-One Matching, it has not yet been shown to be false experimentally. We introduce a novel test to empirically differentiate One-to-One Matching from richer content-based strategies. The crucial scenario is one in which infants hear a sentence with fewer arguments than participants that they represent in the accompanying scene. One-to-One Matching predicts that this sentence should not be a good fit for the scene as they initially represented it. Thematic Linking does not necessarily predict a mismatch, as long as the grammatical relations of the arguments can link in the right way to the particular participant relations that infants perceive the event as having. Here, we pit a transitive clause against a putatively 3-participant event concept. This allows us to focus on the clause type for which children have shown the most reliable behavior in past studies. It also allows us to test an assumption of One-to-One Matching that has not been previously examined: this strategy should hold for any number of arguments and participants, including those greater than two.

The ambiguity in children's previous behavior with intransitive clauses reveals a complication with this approach: this test case will fail to diagnose their bootstrapping strategy if there is indeterminacy in how children readily represent the event stimuli in an experiment. These representations cannot be assumed; in order to be sure about which bootstrapping strategy infants are using, we need an independent measure of the structure of their event representations. In ***Experiment 1***, we introduce a method to norm infants' representations of a scene in which a girl takes a truck from a boy, confirming that they readily view this event as a TAKING with 3 participants. In ***Experiment 2***, we then show that 20-month-olds allow that 3-participant concept to be labeled by a clause with only two arguments: *The girl pimmed the truck*. This mapping between a 3-participant concept and

a 2-argument clause is inconsistent with One-to-One Matching, but is consistent with Thematic Linking. If infants represent *the girl* as subject and *the truck* as object, then they should allow this sentence to describe a 3-participant event in which the girl is agent and the truck is patient, even if the third participant is not mentioned. Thus, this sequence of experiments empirically distinguishes the role of purely structure-based bootstrapping from more flexible alternatives, suggesting that infants privilege the grammatical and thematic relations of arguments above argument number in their early verb learning.

Experiment 1

Experiment 1 adapts a habituation-based method introduced by Wellwood et al. (2015) and He (2015) to diagnose 10- and 11-month-olds' nonlinguistic representations of a "taking" scene. We chose this age in order to test infants before they have begun substantial word or syntax learning, thus minimizing potential effects of the exposure language on their scene percepts. This allows us to identify the likely conceptual representation under which infants perceive events in our experimental task, independent of language.

This step requires a method for isolating the relations that are perceived as participants from all of the other various relations that the event concept entails. If an event is a "taking," then it has an agent of taking, a patient or theme that is taken, a victim or source from which the theme is taken, some manner of transfer, a duration of transfer, a particular location of taking, and so on. But in any particular experience of a taking, its psychological representation may make only some of the entailed relations explicit, foregrounding them as participant relations.

For our "taking" stimulus scene, which relations are privileged in the concept that infants view it under? Perhaps they view it under a 3-participant representation that privileges the agent, patient, and source as participants (3a), or under a 2-participant representation that privileges only the agent and patient (3b), or under a 1-participant representation that privileges only the agent (3c) – and there are many other options (Tatone et al., 2015, 2021).

(3) a. TAKING-AGENT-PATIENT-SOURCE — 3-participant
b. TAKING-AGENT-PATIENT — 2-participant
c. TAKING-AGENT — 1-participant

Each representation in (3) describes a way of mentally encoding a taking event. As such, they all share common entailments: any taking requires an agent, a patient, a source, a location, a time, and so on (see the discussion of the various entailments of SINGING above). These representations thus do not differ in which events in the world they apply to. They only differ in terms of which participants they explicitly encode as psychologically privileged entailments. Such "participant relations" occupy one side of the bootstrapping equation.

To tease apart these possibilities, we use a method introduced in Wellwood et al. (2015) and He (2015), building off earlier work by Gordon (2003). The logic is as follows: if infants treat changes to the hypothesized participant structure under which an event is viewed as more noteworthy than changes to the physical properties of the event, then we might take this as indirect evidence about the structure under which they had viewed the event. He (2015) habituated 10-month-olds to a silent scene in which a girl jimmies open a box using a lever. At test, one group of infants saw the girl now open the box with her hand, with the lever still visible by her side. Another group saw the girl continue to open the box using the lever, but from the right instead of from the left. All jimmyings have a direction of opening, so this also represents a change to one of the entailed relations of the event predicate. However, infants dishabituated (recovered attention) only when the lever was no longer used as an instrument, and not to the change in direction. This pattern was also observed when the lever was added as an instrument at test: infants dishabituated when the girl switched from opening the box with her hand to opening it with the lever, but not when she opened the box with her hand

from the opposite direction. To the extent that infants' differential response to these two types of changes cannot be readily explained by other perceptual factors, this might be taken as indirect evidence for the structure under which they viewed the "jimmying" scene. Specifically, it suggests that they viewed the lever as filling a more psychologically potent relation than the direction of opening, potentially one that is privileged as a participant role in their event representation.

In this experiment, we habituated infants to a silent video in which a girl picks up a toy truck and moves it toward herself, while a boy sits idly by. By hypothesis, this event might be viewed as a 2-participant PICKING-UP, whose participants are the girl and the truck. At test, infants saw one of two changes. One group saw a possible change in participant structure: the boy was now holding onto the truck, and watched as the girl picked it up out of his grasp. By hypothesis, infants might perceive this scene as a 3-participant TAKING, with the boy now a participant. A second group of infants saw a change to the manner of motion: the boy remained uninvolved, but the girl now slid the truck toward herself instead of picking it up. In this case, it is plausibly not the number of participants that differs from habituation to test, but rather the type of motion. If infants dishabituate differentially to the participant change compared to the manner change, then all else equal, this suggests that they view the taking event under a 3-participant concept in which the boy fills a psychologically privileged relation.

Experiment 1a

Method

Participants. Participants included 32 typically-developing infants (15 males, 17 females) from the greater Washington, D.C. area. They had a mean age of 10;22 months (range = 9;16–12;15). 25 additional infants participated but were excluded prior to any analysis due to experimental error (1), parental interference (3), failure to finish the experiment (5), and failure to habituate (16). All infants were recruited online or over the phone through the University of Maryland's Infant and Child Studies Consortium database. Informed parental consent was obtained in accordance with the protocols of the University of Maryland's Institutional Review Board (approval 366,591–15).

Materials. Stimuli consisted of three types of silent live-action videos (see Figure 1), prepared using Adobe Premiere. In PICKING-UP videos, a girl picked up a toy truck from the center of a table and moved it toward herself in the presence of a boy, who was a disinterested bystander. In TAKING videos, the same girl moved the same truck in the same motion, but with boy now looking at and gripping the truck before she reached for it. In SLIDING videos, the girl moved the truck toward herself by sliding it across the table, while the boy again sat idly by, as in the PICKING-UP video. The PICKING-UP and TAKING videos were alike in all respects except for whether the boy was idle or an active participant in the event; the PICKING-UP and SLIDING videos were alike in all respects except for the manner of motion of the truck. A silent video of a butterfly on a flower was additionally prepared as an attention-getter stimulus.

Procedure. After obtaining informed parental consent, infants and parents were led to a room with a 51-inch widescreen television mounted directly below a high-resolution video camera. Stimuli were presented on the TV monitor using the Habit software (Cohen et al., 2004) and an infant's eye gaze was recorded through a live feed from the video camera. Infants were seated 66 inches away from the TV monitor, either on their parent's lap or in a highchair, with parents remaining in the testing room in a chair behind them. Parents were instructed to close their eyes and refrain from interacting with their child, speaking to their child, or pointing to the monitor. In an adjacent room, an experimenter coded an infant's eye gaze by pressing a key whenever the infant attended to the TV monitor, and releasing it whenever the infant looked away. A second experimenter controlled the camera's pan and zoom, to ensure that the infant's face remained in view throughout the duration of the experiment. A video of the infant's face and the corresponding stimuli was recorded using QuickTime.

The experiment used the Habituation-Switch paradigm (Werker et al., 1998; Younger & Cohen, 1985), and was structured as follows. First, the attention-getter stimulus was displayed. Once the infant fixated the attention-getter, the experimenter began the habituation phase. Each trial of the

habituation phase presented up to three different tokens of the PICKING-UP event. Each event lasted 6 seconds, for a maximum trial length of 18 seconds. A habituation trial would end if that maximum trial length was reached or if the infant looked away from the TV monitor for 2 seconds at any point. At the end of each trial, the attention-getter was displayed until the infant oriented toward the television, at which point the next trial would begin.

The habituation phase of the experiment lasted a maximum of 12 trials, with the actual number varying depending on when an infant reached the habituation criterion. We considered this criterion to have been reached when an infant's average looking time within a moving window of 3 trials dropped below 50% of their average looking time during their most-attended 3-trial window (Werker et al., 2002). So, while some participants took all 12 trials to reach the habituation criterion, others habituated after only 4 trials. Infants who did not reach the habituation criterion were excluded from the final sample.

Once an infant was considered habituated, the test phase of the experiment began. This phase always consisted of two trials, each of which contained 3 tokens of a new event type. As during the habituation phase, a test trial ended if the maximum trial length (18 seconds) was reached or if an infant looked away for 2 seconds. The event type during the test trials differed by condition. Infants in the "participant change" condition were exposed to TAKING videos, which differed from the habituation videos in that the boy also participated in the event. Infants in the "manner change" condition were exposed to SLIDING videos, which differed from the habituation videos in that the girl slid the toy truck toward herself instead of picking it up. Infants were randomly assigned to one of the two conditions. Following several other habituation-based tasks (de Carvalho et al., 2019, 2021; He, 2015; He & Lidz, 2017), we use a between-subjects rather than a within-subjects manipulation in order to avoid inducing a comparison between our test videos themselves – a comparison that would be invited by exposing a given infant to both TAKING and SLIDING at test. That is, we are interested in measuring the degree to which infants treat our test videos as distinct from our habituation videos, but not the degree to which they treat the two test videos as distinct from each other. We confirmed that the two groups of infants were matched in age (participant change: $M = 10;25$, range = 9;16–12;15; manner change: $M = 10;19$, range = 9;21–12;1; Welch's $t(29.82) = 0.72$, $p = .48$) and gender (participant change: prop. female = 0.50; manner change: prop. female = 0.56; two-tailed $p = 1$, Fisher's Exact Test). We also confirmed that they did not differ significantly in their habituation profiles (see "Results" below).

Predictions. If infants view our TAKING videos under a concept that differs from the PICKING-UP event in its structure – specifically, under a 3-participant rather than a 2-participant concept – then we predict an asymmetry in infants' dishabituation behavior in the two conditions. The change in the boy's posture and gaze should signal a change in his participanthood; if infants in the participant change condition perceive this change accordingly, then they should dishabituate. Importantly, this should also be perceived as a more substantial change than the change in the girl's manner of motion, because the SLIDING event does not differ from the PICKING-UP event in its participant structure (even though it differs in terms of its content). To the extent that infants dishabituate in the manner change condition, this hypothesis predicts that they should dishabituate to a lesser degree than infants in the participant change condition. However, if infants do not perceive the two types of changes as involving different effects on participant structure, then all else equal, we predict no asymmetry in dishabituation.

Results

To confirm online coding reliability, videos of six experimental sessions were recoded frame-by-frame by a different experimenter. Intercoder agreement was above 87% for all videos.

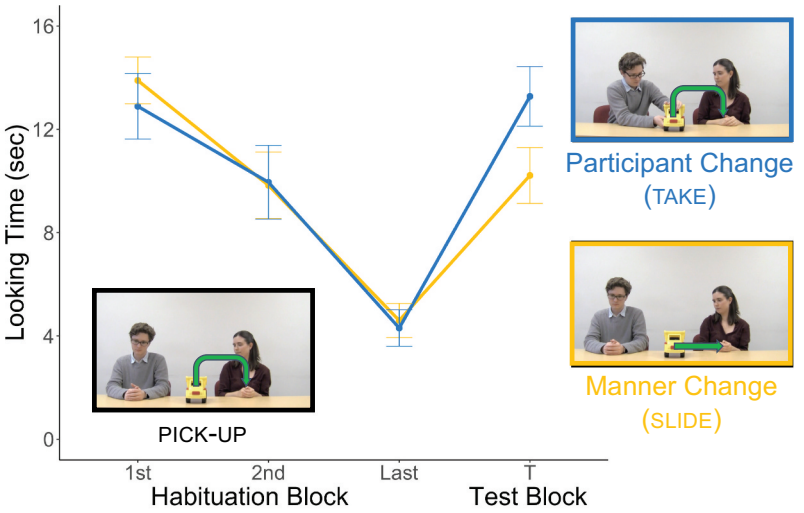

Figure 1. Experiment 1a. Infants' average looking time during the first block of the habituation phase (trials 1 and 2), the second block of the habituation phase (trials 2 and 3), the last block of the habituation phase (trials −2 and −1), and both trials of the test phase (trials T1 and T2). Both groups saw the picking-up video during habituation. Infants in the "participant change" group saw the taking video at test, whereas infants in the "manner change" group saw the sliding video. Error bars represent SEM.

To visualize infants' looking behavior during the habituation and test phases in the experiment, we calculated each infant's average looking time across blocks of two trials (trial 1 and 2, trial 3 and 4, etc.). Figure 1 displays average looking times by trial block across all infants within each condition.

We first confirmed that infants in both groups displayed no significant differences in their habituation profiles. Infants in the participant change condition habituated after an average of 7.81 trials (SD: 2.88, range: 4–12), and infants in the manner change condition habituated after an average of 8.31 trials (SD: 2.75, range: 4–12). This difference was not significant (Welch's $t(29.94) = 0.50$, $p = .62$). We further compared the first and last blocks of the habituation phase in a linear mixed-effects regression, using the *lme4* package in R (Bates et al., 2015). The model included fixed effects of condition (participant change vs. manner change), habituation block (trial 1 and 2 vs. trial −2 and −1), and their interaction, and a random intercept for subject. Our dependent measure was looking time in seconds for each trial. For this and all following analyses, factor contrasts were sum-coded and significance testing was performed through likelihood ratio tests. Fixed effects were examined by comparing a model containing the effect against a model that differed only in that it lacked the relevant effect. We find a significant main effect of block ($\chi^2(1) = 76.52$, $p < .001$) but no main effect of condition ($\chi^2(1) = 0.37$, $p = .54$) and no interaction between the two ($\chi^2(1) = 0.18$, $p = .67$). That is, condition had no effect in the habituation phase, during which infants in both groups were exposed to the same stimulus.

To determine the extent to which infants treated the test videos as distinct from the habituation videos, we compared the final two trials of the habituation phase and the two trials of the test phase in a mixed effects regression with fixed effects of condition (participant change vs. manner change), phase (habituation vs. test), and their interaction, and a random intercept for subject. We find a significant main effect of phase ($\chi^2(1) = 58.99$, $p < .001$), indicating that infants dishabituated upon reaching the test phase. Importantly, we find a significant interaction between condition and phase ($\chi^2(1) = 4.35$, $p < .05$), indicating that infants dishabituated to different degrees when shown the participant change video (TAKING) than when shown the manner change video (SLIDING) at test. This is further confirmed by a post-hoc comparison of looking times during the test trials in each condition. Infants in the participant change condition looked longer at test than infants in the manner change condition ($t(70.8) = 2.37$, $p < .05$, Satterthwaite's approximation for degrees of freedom).

Experiment 1b: Control

In order to take these observed differences in dishabituation as indirect evidence for differences in infants' conceptual representations, we must rule out two alternative explanations. The first is that our test stimuli may not be equally interesting to infants, independent of the habituation phase. If infants who had not been habituated to PICKING-UP were still to find our TAKING video more interesting than our SLIDING video, then differences in looking times to these videos would not be informative about their treatment of the participant vs. manner change. Second, if there are larger perceptual differences between our TAKING and PICKING-UP videos and smaller perceptual differences between SLIDING and PICKING-UP, then infants' asymmetrical dishabituation behavior might be explained by their reactions to those perceptual asymmetries. We conducted a control experiment to test both of these alternative accounts.

Method

Participants. Participants included a second sample of 32 typically-developing infants (13 males, 19 females) recruited from the greater Los Angeles area. They had a mean age of 11;7 months (range: 9;20–12;17). 29 additional infants participated but were excluded prior to any analysis due to experimental error (12), equipment malfunction (2), parental interference (2), and failure to habituate (13). All infants were recruited online or over the phone through the University of Los Angeles (UCLA) Developmental Subject database. Informed parental consent was obtained in accordance with the protocols of UCLA's Institutional Review Board (approval 10–001562).

Materials and Procedure. Significant care was taken to match testing conditions at our UCLA location to those in Experiment 1a. Stimulus presentation and coding were conducted through the Habit program using the same testing and coding protocols as above, except for the following differences. Stimuli were displayed on a 46-inch television, with infants seated 42 inches away to match their viewing angle as closely as possible to that in Experiment 1a. The video feed for the experimenter in an adjacent room was provided by a high-definition video camera located directly below rather than above the TV monitor.

The experiment had two parts. In Part A, we measured infants' independent interest in our test stimuli by presenting them with the exact test phase from Experiment 1a, without the habituation phase. One group of infants saw the two TAKING trials from the original participant-change condition. A second group saw the two SLIDING trials from the original manner-change condition. Infants were randomly assigned to one of the two groups. Just as before, each trial ended if the maximum trial length was reached or if an infant looked away for 2 seconds.

In Part B, we tested the possibility that there were purely perceptual asymmetries between our original habituation and test stimuli. To do this, we repeated Experiment 1a with the stimuli displayed upside-down. This manipulation holds constant the physical differences between our habituation and test scenes – including object and actor position, motion, and relative size on the screen – while making the events more difficult to conceptualize. This follows a control manipulation in Gordon's (2003) similar habituation-based design, and the logic is similar to speech perception tasks that control for low-level acoustic differences by playing speech stimuli in reverse (Dehaene-Lambertz et al., 2002; Marno et al., 2015; Mehler et al., 1988; Peña et al., 2003; Vouloumanos & Werker, 2004, 2007). All materials and procedures were identical to those in Experiment 1a except for the 180° rotation of the stimuli. Infants in the participant change condition were habituated to our upside-down PICKING-UP videos and tested on our upside-down TAKING videos; infants in the manner change condition were habituated to our upside-down PICKING-UP videos and tested on our upside-down SLIDING videos. Conditions were assigned relative to Part A, such that no infant saw the same video upside-down

that they had earlier viewed right-side-up. Between Parts A and B, infants viewed a 30-second video of moving toys accompanied by music to re-set their attention.

We confirmed that both groups of infants were matched in age (participant change: $M = 11;13$, range = 10;0–12;17; manner change: $M = 11;1$, range = 9;20–12;12; Welch's $t(30.00) = 1.04$, $p = .31$) and gender (participant change: prop. female = 0.56; manner change: prop. female = 0.63; two-tailed $p = 1.00$, Fisher's Exact Test). We further confirmed that they displayed no differences in their habituation profiles in Part B (see "Results" below).

Results

To confirm online coding reliability, videos of six experimental sessions were again recoded frame-by-frame by a different experimenter. Intercoder agreement was above 86% for all videos.

To determine whether infants found our TAKING and SLIDING videos equally interesting without a habituation phase, we first calculated each infant's average looking time during the two Part A trials. Figure 2a displays average looking times across all infants within each condition. Unlike in Experiment 1a, we found no significant differences in looking times between conditions (Welch's $t(27.76) = 0.53$, $p = .60$). That is, when infants viewed these videos without prior habituation to PICKING-UP, they did not find TAKING to be more interesting than SLIDING.

We then analyzed infants' habituation profiles to the upside-down stimuli in Part B. Figure 2b displays average looking times across blocks of two trials for each condition. We again confirmed that infants in both groups displayed no significant differences in their habituation profiles. Infants in the participant change condition habituated after an average of 6.69 trials (SD: 1.92, range: 4–10), and infants in the manner change condition habituated after an average of 7.13 trials (SD: 2.47, range: 4–12). This difference was not significant (Welch's $t(28.28) = 0.56$, $p = .58$). We then compared looking times during the first and last blocks of the habituation phase across conditions, using the same linear mixed-effects regression model as in Experiment 1a. We find a significant main effect of block ($\chi^2(1) = 98.58$, $p < .001$) but no main effect of condition ($\chi^2(1) = 0.11$, $p = .74$) and no interaction between the two ($\chi^2(1) = 0.26$, $p = .61$). That is, infants in both conditions displayed the same habituation behavior.

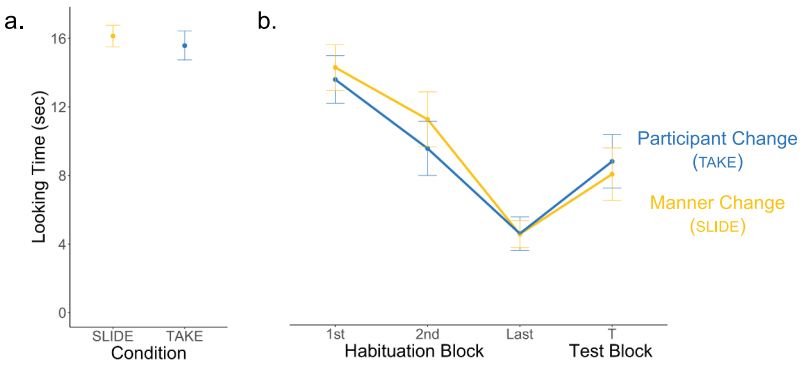

Figure 2. Experiment 1b. a. Infants' average looking times to the taking and sliding videos in Part A, with no prior habituation phase. b. Infants' average looking times in Part B during the first block of the habituation phase (trials 1 and 2), the second block of the habituation phase (trials 2 and 3), the last block of the habituation phase (trials −2 and −1), and both trials of the test phase (trials T1 and T2). Both groups saw the upside-down picking-up video during habituation. Infants in the "participant change" group saw the upside-down taking video at test, whereas infants in the "manner change" group saw the upside-down sliding video. Error bars represent SEM.

To compare dishabituation patterns between conditions, we analyzed the final two trials of the habituation phase and the two trials of the test phase, using the same linear mixed-effects regression model as in Experiment 1a. Just as in Experiment 1a, we find a significant main effect of phase ($\chi^2(1) = 21.68$, $p < .001$), indicating that infants dishabituated upon reaching the test phase. But unlike in Experiment 1a, we find no significant interaction between condition and phase ($\chi^2(1) = 0.21$, $p = .65$), indicating that infants dishabituated to the same extent when shown the upside-down participant change video (TAKING) compared to the upside-down manner change video (SLIDING). Post-hoc comparisons confirm no significant differences in looking times during the test trials between conditions ($t(60.3) = 0.53$, $p = .60$, Satterthwaite's approximation for degrees of freedom).

Discussion

In Experiment 1a, we find that infants registered a change from an event in which a girl moves a toy truck toward herself with a boy sitting idly by, to an event in which a girl takes the truck from the boy's grasp – by hypothesis, a change from a 2-participant to a 3-participant event representation. They moreover dishabituated to a greater extent than when presented with change in the manner of motion. In Experiment 1b, we find that this asymmetry is not explained by independent differences in interest in our test scenes, or by purely perceptual asymmetries between conditions. When infants viewed the test scenes without a prior habituation phase, they found both equally interesting. And when the stimuli were presented upside-down, preserving physical differences but obscuring conceptual differences, infants showed no asymmetrical pattern of dishabituation. This suggests that infants' behavior in Experiment 1a reflects their reactions to the conceptual changes in this experiment: they viewed the boy's involvement in the event as more noteworthy than a novel manner of motion.

To be sure, we also found that infants distinguished the girl picking up the truck from the girl sliding the truck. Not every change that infants can detect suggests a difference in the structure of the representations deployed; we expect that infants can distinguish events that differ in their content, even if they are perceived as having the same number of participants. The important finding for our purposes is that infants reacted more strongly to the change in the boy's involvement. They did so even though our control task suggests that the physical difference between PICKING-UP and SLIDING was perceived as equal to the physical difference between PICKING-UP and TAKING. If the dishabituation observed between PICKING-UP and SLIDING reflects these physical differences, along with a difference in the content of the event representations, then what explains the additional dishabituation observed between PICKING-UP and TAKING? Here we follow a similar argument from a seminal study by Leslie and Keeble (1987): to the extent that this additional dishabituation is not readily attributable to perceptual factors, then we argue that it supports a difference in conceptual representations. In particular, we argue that it supports a difference in the participant structure of the concepts, a change from a 2-participant to a 3-participant representation. That is, infants' behavior receives an explanation if they readily view the TAKING scene with the boy filling an explicitly-represented participant role, alongside the girl and the truck. This behavior is not predicted if they view the TAKING scene under the same conceptual structure as the other events, with only the girl and the truck as participants. We might thus take this finding as evidence, albeit indirect, for a 3-participant representation:

(4) TAKING-AGENT-PATIENT-SOURCE

This is not to say that all scenes that adults can label as "takings" are viewed under a 3-participant concept. If a source is particularly unresponsive, it may be ignored. For example, Tatone et al. (2015) found that 12-month-olds failed to differentiate scenes in which a personified square took an apple from a personified circle, from scenes in which a square obtained an apple while the circle watched (similar to our "picking-up" scenes). On the other hand, infants the same age successfully

differentiated similar scenes in which a square gave an apple to a circle, from scenes in which a square discarded an apple while a circle watched. This suggests an asymmetry in the extent to which the recipient or source relation is highlighted when representing GIVINGs versus TAKINGs (see also Yin et al. (2020), who argue for a similar GIVING/TAKING asymmetry in adults). But as Tatone and Csibra (2020) show, adding even a small social cue can cause infants to represent cartoon scenes under a 3-participant TAKING concept. For example, if the actors look at each other before the apple is moved, infants treat the "taking" scenes on par with the "giving" scenes, representing both under concepts that make three participants explicit. In our "taking" event, the boy retracts his hands and follows the truck with his eyes as it is moved out of his grasp. These subtle reactions may serve as the type of social cues that encourage infants to represent the scene under a 3-participant concept (for related findings suggesting that similar postural and social cues can affect the perception of participanthood, see Hafri et al., 2013; Papeo et al., 2017).

Additional work might take further steps to explore these issues. But having gained some initial evidence for infants' representation of this particular "taking" scene, we are in a better position to identify which principles they use when mapping a sentence to this scene, at the age when they are learning verbs. Namely, do they expect that clause arguments must match one-to-one the event participants that they perceive, or do they more flexibly link particular grammatical and participant relations? In Experiment 2, we pit this "taking" stimulus scene against a sentence with fewer than three arguments, and we ask what inferences 20-month-olds will draw about the meaning of the verb in that sentence. This provides a test case for differentiating One-to-One Matching from more specific content-based bootstrapping inferences.

Experiment 2

Experiment 2 tested whether 20-month-old English learners allow a 2-argument clause to describe our TAKING scene as they readily perceive it, under a 3-participant concept. We chose this age range as it represents the youngest age most commonly tested in earlier experiments on infants' verb learning. We used a novel verb learning task adapted from the Verb Extension paradigm (Waxman et al., 2009). Infants were familiarized to a version of the "taking" stimulus scene (a girl takes a toy truck from the boy), described by a novel verb in a transitive clause: *The girl pimmed the truck*. On the basis of these familiarization trials, infants should make an inference about what kinds of events *pimmings* are. We then tested which inference they made by asking what else counted as an instance of *pimming* for them. At test, infants were prompted to find *pimming* in the context of two candidate videos. One showed the girl still taking the truck from the boy (another token of the "taking" scene). The second showed the girl moving the truck toward herself in the same way, but without the boy present (a "grabbing" scene). Measuring infants' looking preferences to these two videos allows us to determine whether they concluded that transitive *pimmings* could be 3-participant TAKINGS, or whether they concluded that transitive *pimming* must label a 2-participant construal of the scene, or a 2-participant sub-event, involving only the girl and the truck.

To control for the possibility that infants' preferences at test may not reflect inferences based on syntax, but instead a general bias to map a novel verb to the familiar scene, we compared their behavior in this experimental condition against a control condition. Infants in the control group saw identical video stimuli, but were familiarized to the novel verb in an intransitive clause: *The truck pimmed*. Both One-to-One Matching and Thematic Linking predict that this sentence should not be perceived as a good fit for a 3-participant TAKING, but more likely describes some aspect of the truck's motion, which is the same in both test videos. Under the One-to-One strategy, this clause must describe an event construal or a salient sub-event with the truck as the sole participant. Under the Thematic Linking strategy, it likely describes the truck as the patient of a change, potentially in the absence of an agent. Because the clause syntax gives no reason for infants to prefer the "taking" scene over the "grabbing"

scene at test, we can use any residual preferences in this condition as a baseline for infants' general familiarity biases in this design.

Method

Participants

Participants included 48 typically-developing infants (25 males, 23 females) from the greater Washington, D.C. area. They had a mean age of 20;5 months (range: 19;0–21;28). Participants were recruited with the criterion that they heard English during at least 80% of their waking hours. An additional 9 infants were tested but excluded prior to any analysis due to inattentiveness (6), equipment malfunction (1), or less than 80% English exposure (2). All infants were recruited online or over the phone through the University of Maryland's Infant and Child Studies Consortium database. Informed parental consent was obtained in accordance with the protocols of the University of Maryland's Institutional Review Board (approval 366,591–15).

Participants' total productive vocabulary was collected by parental report using the Words and Sentence MacArthur-Bates Communicative Development Inventory (MCDI) (Fenson et al., 1993). Mean total words produced were 133.65 (range: 4–591); mean total verbs produced were 16.85 (range: 0–118).

Materials

Visual stimuli consisted of live-action video of two actors performing actions with inanimate objects. Three event types (SHAKING, OPENING, HUGGING) were used in training trials, and one event type (TAKING) was used during the experimental trial. Six different tokens of each event type were filmed and edited in Adobe Premiere to create the trial structure in Table 1. Tokens were edited to be 7.5 seconds in duration during the familiarization and contrast phase and 5.5 seconds during the test phase. The TAKING videos were designed to be nearly identical to those used in Experiment 1, with one difference: the girl takes the truck from the boy by sliding it across the table instead of lifting it, as adult piloting found the sliding motion to be more natural.

An additional eight videos of different events were created: four that were used in the "contrast" phase of each trial, as a negative exemplar of the verb's event, and four that were paired with a token of the familiarization videos during the preferential looking phase. The negative contrast video involved the same actors, but introduced a new action, with a different manner of motion, performed on a new object. The pairs of videos created for the preferential looking phase of the training trials contrasted different actions (e.g. SHAKING vs. SPINNING) with the same actors and objects. For the experimental trial (TAKING vs. GRABBING), the manner of motion was held constant in the two videos but the boy was only present in one of them. Unlike in Experiment 1, we removed the boy as a bystander from the GRABBING video, so as to highlight better the contrast between the events when presented side-by-side. However, we

Table 1. Structure of the experimental trial (taking), Experiment 2. Sample audio is for the experimental condition; the control group saw identical video but heard intransitive clauses during all trial phases.

		Sample Audio (Experimental Condition)	Video	
Familiarization (30s)		*Look, the girl is gonna pim the truck! She just pimmed the truck!*	Girl takes truck from boy	
Contrast (15s)	Negative	*Uh-oh, she's not gonna pim THAT. She didn't pim THAT.*	Girl pokes tower held by boy	
	Positive	*Yay, she's gonna pim the truck! She pimmed the truck!*	Girl takes truck from boy	
Test (16.5s)	Baseline	*Now look, they're different!*	Girl takes truck from boy	Girl grabs truck, no boy
	Response	*Find the one where she's pimming the truck. Where is she pimming the truck?*		

note that prior preferential looking tasks find that the presence or absence of a bystander to an event does not affect children's inferences about novel verb meaning (Yuan et al., 2012). See Table 1 for a description of the videos used during the experimental trial, and the Appendix for the full list of videos used in the training trials.

Audio stimuli were recorded by a female native speaker of American English using child-directed speech. Stimuli were edited in Adobe Audition and Praat, and combined with video stimuli in Adobe Premiere. During familiarization, audio was timed to frame each action: a future-tense sentence (e.g., *The girl is going to pim the truck!*) ended as the action began, and a past-tense sentence (e.g., *The girl just pimmed the truck!*) began as soon as the action ended. At test, sentence onset was timed to coincide with the beginning of each looped video. Stimuli in the two conditions were identical except for the syntactic frame used during the experimental trial: the verb *pim* was presented in a transitive frame in the experimental condition and in an intransitive frame in the control condition. A complete list of sentences used in both conditions is provided in the Appendix.

Procedure

The same parental consent and testing protocols were followed as in Experiment 1, with the exception that infants' looking fixations were not live-coded during the experimental session. Instead, a video of the infant's face and the corresponding stimuli was recorded using QuickTime, and an experimenter in an adjacent room controlled the camera's pan and zoom to ensure that the infant's face remained in frame throughout. Each experiment lasted 5.6 minutes.

Infants were randomly assigned to either the experimental or the control condition. In each condition, the experiment followed the same structure. Infants were first introduced to the two actors, who appeared on different sides of a black screen for 7 seconds each, waving and smiling. They were introduced as *the boy* and *the girl*, and were also referred to by pronouns (e.g. *Look, it's a girl! Do you see her? There she is!*). They then appeared for 15 seconds in split-screen, and infants were prompted to find each one in turn.

After actor introductions, infants saw four trials that each followed the structure in Table 1, adapted from Waxman et al. (2009). During the familiarization phase, infants saw four video tokens of a particular event type appearing on different sides of the screen, described twice by a verb in a full sentence. During the contrast phase, infants saw a new video that was described in downcast intonation as a negative example of the verb's event, followed by another token of the familiarization scene, described in upbeat intonation as a positive example. This phase was included to facilitate infants' recognition that our novel verb *pim* has a specific meaning (Waxman et al., 2009). During the test phase, two videos were presented concurrently on different sides of the screen: another token of the familiarization scene (e.g. the girl takes the truck from the boy) and a new scene (e.g. the girl grabs the truck, without the boy). Participants were randomly assigned to different lists to counterbalance the screen side of the familiar vs. new video, and whether the familiar video matched the side on which the positive contrast video had appeared. The two test videos were first accompanied by uninformative audio (*Now look, they're different!*), providing a baseline measure of differences in salience. The videos then played on loop twice more, and infants were prompted to find the verb's event. Finally, the videos disappeared to a black screen, and a new trial began. To focus infants' attention, trials were interleaved with either a 4-second still image of a baby face with audio of a baby giggling, or a 14-second video of moving toys accompanied by music.

The first three trials consisted of training trials with known verbs, which were included to facilitate infants' familiarity with the experimental procedure before the novel verb was introduced in the fourth trial (Scott & Fisher, 2009; Yuan & Fisher, 2009). So as not to bias infants toward any particular syntactic frame, the three training trials each presented a verb in a different argument structure: one ditransitive, one transitive, and one intransitive frame, with order counterbalanced across participants.

We chose the verbs *give*, *shake*, and *open*, which rank high in familiarity for 20-month-olds (Frank et al., 2016). Each training trial had the same structure as the experimental trial in Table 1. See the Appendix for the full list of trials.

Predictions

If 20-month-olds primarily rely on One-to-One Matching for bootstrapping verb meanings, then we predict that infants in both our experimental and control conditions will perceive a mismatch between the syntax of the familiarization sentence and the "taking" stimulus scene, as they naturally perceive it. When presented with *pim* in a 2-argument or 1-argument clause, infants should conclude that this clause cannot describe an event that they readily view under a 3-participant concept; instead, it must be describing an event or sub-event that has fewer participants. In the experimental condition, infants who hear *The girl pimmed the truck* should infer that it describes an event with only the girl and truck as participants, such as the girl's grabbing of the truck. *Pimmings* are not 3-participant TAKINGS, but more likely 2-participant GRABBINGS. In the control condition, infants who hear *The truck pimmed* should infer that it describes an event with only the truck as a participant, such as the truck's motion across the table. *Pimmings* are not 3-participant TAKINGS, but more likely 1-participant MOVINGS. Because the girl moves the truck toward herself in the same way in both test videos, we predict no above-baseline preferences for one video over the other in either condition. To the extent that infants' baseline preferences are affected by a bias for the familiar video, that bias will manifest in the same way for both sentence types. Thus, the One-to-One hypothesis predicts no difference between conditions at test.

Under the hypothesis that infants at this age are instead primarily linking particular grammatical and participant relations, with no expectation of one-to-one matching, we predict that infants will perceive the fit between the familiarization sentences and scenes differently. In the experimental condition, infants who hear *The girl pimmed the truck* should infer that it describes an event that they represent as having a girl as agent and a truck as patient. This sentence can therefore label the 3-participant TAKING concept under which they readily view the familiarization scene, provided that they view the girl and the truck as filling those respective participant roles. *Pimmings* in this context might be 3-participant TAKINGS: the syntax gives no reason to conclude otherwise. However, infants in the control condition should not draw the same conclusion if they are attending to the syntax of the clause and the thematic relations of the clause arguments. Hearing an intransitive clause whose subject names the patient of an event— *The truck pimmed*— should lead infants to infer that it is describing a change to that patient, potentially independent of an agent. That is, the sentence is describing the scene from the truck's perspective, possibly focusing on some aspect of its motion. *Pimmings* in this context are not likely to be 3-participant TAKINGS; if anything, they are some type of MOVING. This inference depends on using the thematic content of the intransitive subject and not merely the number of arguments in the clause, even in the weaker fashion that we had called Unidirectional Mapping. If infants only map the number of clause arguments unidirectionally to the perceived number of event participants, without any sensitivity to thematic content, then they should allow an intransitive clause whose sole argument is *the truck* to label any event with a truck as participant: *pimmings* in this case might easily be 3-participant TAKINGS, just as in the experimental condition. Thus, only if infants are indeed using grammatical and thematic content above and beyond argument number in this task do we predict a difference between the two conditions at test: infants will show a greater preference for the TAKING video in the experimental condition than in the control condition.

Results

Videotaped recordings of the experimental test phase were coded using EyeCoder software (Fernald et al., 2008). An experimenter advanced each muted video frame-by-frame and coded whether the

infant was looking at the left or right side of the screen, or neither. Data were coded by two experimenters, with intercoder reliability above 90% (Cohen's Kappa > 0.90).

Two windows of analysis were selected within the test trial. The baseline window spans the 3 seconds before the offset of the novel verb *pimming*, and the test window spans the 3 seconds after novel verb offset. This allows us to measure any baseline preferences for one of the two videos, and examine how these preferences shift after infants are asked to find *pimming*. At each frame, we calculated whether an infant was looking at the TAKING video, the GRABBING video, or neither. For purposes of visualization, we then calculated infants' average proportion of time spent looking to the TAKING video during each window of analysis, out of time spent looking toward either video. These average proportions of looks to TAKING are plotted by condition and window in Figure 3. Visual inspection reveals no overall preference for either video during the baseline window, but preferences that differ by condition during the response window: infants in the experimental condition show a strong preference for the TAKING video, whereas infants in the control condition do not.

To assess the reliability of this pattern, we conducted a binomial logistic mixed effects regression with total frames spent looking to the TAKING video and to the GRABBING video as the dependent variable. Note that this measure is no longer a proportion. Although we present proportions of looks to TAKING in Figure 3 for ease of visualization, our analysis instead uses the raw counts of frames that each infant spent looking to TAKING and the counts of frames spent looking to GRABBING, order to retain information about how much total data an individual infant is contributing (Lidz et al., 2017).[3] Fixed effects included window, condition, and their interaction. In order to account for individual differences in subject preferences, our full model also included a random intercept for subject and a random slope for window. Likelihood ratio tests revealed a significant main effect of window ($\chi^2(1) = 5.32$, $p < .05$), but no significant main effect of condition ($\chi^2(1) = 0.69$, $p = .40$). Importantly, consistent with the visual pattern observed in Figure 3, we find a significant 2-way interaction of window and condition ($\chi^2(1) = 3.98$, $p < .05$). Pairwise post-hoc comparisons using Z-tests, corrected with

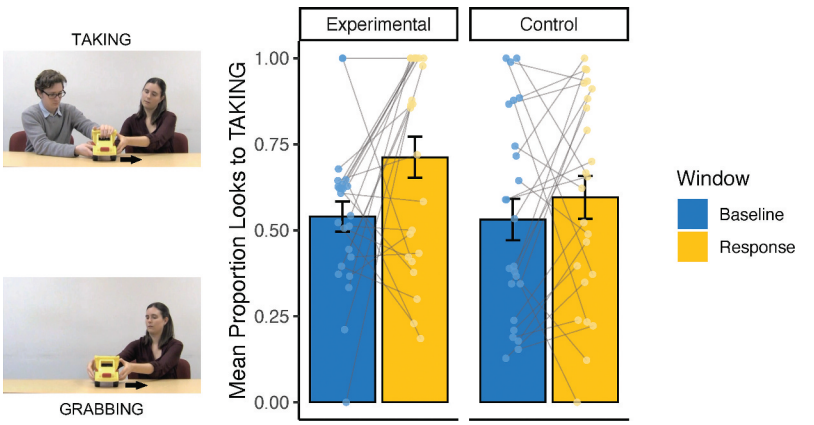

Figure 3. 20-month-olds' average proportion time spent looking to the taking video vs. the grabbing video during the three seconds prior to novel verb offset (the baseline window) and the three seconds following novel verb offset (the response window). Infants in the experimental condition heard the novel verb in a transitive frame, and infants in the control condition heard the novel verb in an intransitive frame. Error bars represent SEM. Points show looking proportions for individual infants, with lines connecting baseline and response proportions within-subjects.

[3]A binomial logistic regression is appropriate for modeling counts of looks to the TAKING vs. GRABBING videos, whose sum equals the total looks to the screen in a trial for a given child. With this dependent measure, we are asking about the probability of observing k looks to TAKING out of n looks total in the trial; this follows a binomial distribution with n trials and probability q. We refer the reader to Lidz et al. (2017) for further details.

Holm's sequential Bonferroni procedure, revealed that infants in the experimental condition looked significantly longer to the TAKING video during the response window compared to the baseline window ($Z = -3.07$, $p < .05$). In the control condition, there was no significant difference between the baseline and response windows ($Z = -0.34$, $p = .99$). Thus, infants behaved differently in the two conditions when asked to find the event of *pimming*. Infants who were familiarized to *pimming* in a transitive frame showed an above-baseline preference for the 3-participant TAKING video at test; infants who had been familiarized to *pimming* in an intransitive frame were indifferent, and showed no above-baseline preference for 3-participant TAKING vs. 2-participant GRABBING.

We further examined whether infants' verb vocabulary affected their behavior in this task by fitting another binomial logistic mixed-effects regression model, adding a fixed effect of log-transformed total verbs as reported on the MCDI, and its interaction with window and condition. A model with a random slope for window failed to converge, so the full model included only a random intercept for subject. Likelihood ratio tests again revealed a significant main effect of window ($\chi^2(1) = 96.51$, $p < .001$) and a significant interaction of window and condition ($\chi^2(1) = 45.24$, $p < .001$). The 2-way interaction of log verb vocabulary and window was also significant ($\chi^2(1) = 44.13$, $p < .001$), but no other significant interactions or main effects were found (all ps >0.13). Importantly, there was no significant 3-way interaction of log verb vocabulary, window, and condition ($\chi^2(1) = 1.64$, $p = .20$). These results indicate that infants' prior verb knowledge may have affected the magnitude of their shift in preference from the baseline to the response window, with a lower magnitude for infants with higher verb vocabulary, independent of condition. However, prior verb knowledge did not affect the different patterns of preference that we observe in the two conditions. These different preference patterns were exhibited to the same extent in infants with low and high productive verb vocabulary.

Discussion

In Experiment 2, we found that 20-month-olds did not appear to register a mismatch between a 2-argument clause (*The girl pimmed the truck*) and an event that they readily perceive under a 3-participant concept (a girl taking a truck from a boy). When asked to find another instance of *pimming* at test, infants looked toward another token of the 3-participant TAKING event, preferring this scene over a 2-participant alternative in which the girl moved the truck toward herself in the same way, without the boy. This behavior is not predicted by One-to-One Matching: infants using this strategy should conclude that a 2-argument clause describes a 2-participant construal of the scene. Thus, these results are inconsistent with the hypothesis that infants expect the arguments in a clause to match perceived event participants one-to-one.

Infants' behavior in the control condition rules out the possibility that their preferences resulted from a familiarity bias, rather than sensitivity to clause syntax. If infants had ignored clause syntax in this task, then this predicts that they would map *pimming* to the 3-participant TAKING—the scene present during the familiarization phaseregardless of whether they heard a transitive or intransitive clause. However, we found different behavior in the two conditions: unlike infants who heard transitive syntax, infants who heard intransitive syntax were indifferent between TAKING and GRABBING at test, suggesting that they viewed the GRABBING scene as an equally good instance of *pimming*. This indicates that infants were using the syntax of the familiarization clause when drawing inferences about the novel verb. And this difference between conditions was not predicted by infants' productive verb vocabulary, suggesting that prior verb knowledge did not play an important role in their bootstrapping inferences in this task.

Our findings are not consistent with One-to-One Matching, but they are consistent with Thematic Linking. On this account, infants who hear *The girl pimmed the truck* during familiarization would expect the subject to name the agent and the object to name the patient of the TAKING scene, as they readily perceive it. Because these relations align in the right way, they could then infer that the transitive clause describes the entire 3-participant concept under which they readily viewed the TAKING scene. The sentence would not necessarily push them toward another 2-participant concept: if infants

treat their conceptual percept as equally informative about clause meaning in this context, then they would conclude that *pimmings* can be 3-participant TAKINGS, and most likely are. That is, because the syntax of the clause aligns well with the bootstrapper's most readily accessible conceptual representation, no adjustment of that representation is needed to identify which event the sentence is describing. However, the intransitive sentence (*The truck pimmed*) does not align easily with this 3-participant representation. Given that the inanimate intransitive subject is more likely a patient than an agent, this sentence more plausibly describes some aspect of the truck's motion than the girl's taking of the truck from the boy.[4] Infants who hear this sentence should conclude, given what happens in the scene, that *pimmings* are likely some form of MOVING, predicting no above-baseline preference for either video at test. The current results confirm both of these predictions, providing support for Thematic Linking as infants' bootstrapping mechanism at 20 months.

There may be other ways of accounting for infants' indifference in the intransitive condition. For instance, it is possible that infants at this age only know that this type of intransitive clause cannot describe a 3-participant TAKING specifically, without knowing what other type of events it might describe. If this is the case, the resulting confusion at test might manifest as apparent indifference. Although prior experimental work shows that 2-year-olds are able to draw inferences about the meanings of verbs in different intransitive frames (Bunger & Lidz, 2004, 2008; Scott & Fisher, 2009), further investigation is needed to assess whether 20-month-olds likewise have this ability, and can deploy it in our task. What is important for our purposes is that the 20-month-olds in our task treated transitive and intransitive syntax differently when drawing inferences about verb meaning, and that they used this syntax in a more flexible way than predicted under One-to-One Matching. And, the difference we find between conditions rules out the possibility that infants' inferences reduced entirely to the assumption that the number of arguments in a clause maps unidirectionally to the number of perceived event participants (Unidirectional Mapping), without taking into account the thematic content of those arguments. Because both the transitive and intransitive sentences name participants in the 3-participant TAKING, this weaker number-based strategy predicts that both should easily describe that event as it is readily perceived, resulting in no difference in behavior between the two conditions. Infants' different inferences for transitive and intransitive sentences therefore suggest that they made use of thematic content, and not merely argument number, to constrain their mapping between syntax and meaning.

This argument relies on an important assumption: that 20-month-olds readily perceived the TAKING scene under the same 3-participant structure as the 10-month-olds in Experiment 1, rather than under a 2-participant structure. Our first experiment tested a population with less linguistic experience in order to better diagnose event concepts as they arise in early cognitive development, abstracting away from accidental influences of a child's native language. Although it is possible for infants' event representations to change between 10 and 20 months, it is plausible that older infants would perceive these events under structures that are no less rich than those of younger infants.

These results tell us about a specific step in infants' verb learning: the initial mapping between clause syntax and a particular event that the clause describes, which is the first step in narrowing down the meaning of a new verb. These results do not tell us how infants might generalize the meaning of the verb *pim* beyond the current task, or how they might acquire the meaning of the English verbs *take* or *grab*. Beyond the initial step of pairing a single sentence with a single scene, a learner must generalize across multiple instances of scene-sentence pairings in order to identify the correct conceptual category that a verb picks out (Arunachalam & Waxman, 2011, 2015; Scott & Fisher, 2012; Waxman et al., 2009). For instance, in order to acquire the meaning of the English verb *grab*, a child may need to experience this verb uttered in the context of scenes that are more readily perceived as 2-participant GRABBINGS rather than 3-participant TAKINGS. And to the extent that the

[4]This inference relies on the fact that the intransitive subject is inanimate. An intransitive clause with an animate subject (i.e., *The girl pimmed*) would not likely lead to the same inference. This invites the question of how 20-month-olds would behave with this type of intransitive clause, a question which we leave for future work.

English verb *take* may in fact sometimes refer to events with only 2 participants (those that we have labeled as "grabbings" or "pickings-up"), a child would likewise need to experience contexts that more readily support those event construals, in order to generalize the verb meaning appropriately.

How learners succeed in learning the meanings of verbs in their totality is an important question, but one that is beyond the scope of the current study. Here, our focus is on the principles that guide the very early steps of this process, whereby the syntax of a clause helps the learner identify a single event out of many that might be in the extension of a new verb's meaning. The current results demonstrate that at 20 months, this bootstrapping step is not primarily driven by the expectation of one-to-one matching between clause arguments and event participants. Instead, infants in our task appeared able to use a more flexible strategy to relate their sentence and scene representations, one that suggests sensitivity to differences in thematic content. They allowed a 3-participant event to be described by a transitive clause whose subject names the agent and whose object names the patient, but not by an intransitive clause with an inanimate subject. Infants' behavior was moreover consistent across varying degrees of prior verb knowledge, suggesting that this more sophisticated strategy can be deployed even by infants at early stages of verb learning. Thus, the current findings are consistent with the hypothesis that even very immature learners rely on finer-grained information above and beyond the number of clause arguments at initial stages of syntactic bootstrapping.

General discussion

In the years since the syntactic bootstrapping hypothesis was first introduced (Landau & Gleitman, 1985), a great deal of evidence has accumulated demonstrating (a) that correlations between verb meaning and syntactic distribution are present in speech to children (Fisher et al., 1991; White et al., 2018) and (b) that both adults and children can use this information in acquiring word meanings (Fisher & Song, 2006; Gillette et al., 1999; Piccin & Waxman, 2007; inter alia). But it is not yet certain which correlated properties are in fact represented by learners in their second year of life and used to infer the meanings of action verbs.

Some prior authors suppose that one basis for these inferences, in the second and third years of life, lies in the presumption of a purely structural similarity between syntactic and conceptual representations: the two are assumed to have the same number of variables. But we argue against this view, and in favor of one based on content at both levels. Specifically, we argue that, even in toddlers, bootstrapping inferences are guided not by a presumption of numerical matching, but by knowledge of which syntactic relations go with which thematic relations – or more neutrally, of how a certain asymmetry among arguments in syntax relates to a certain asymmetry among participant relations to an event.

Doing so required a methodological advance over previous work on syntactic bootstrapping. To the extent that children engage in syntactic bootstrapping of verb meaning in natural conditions, they do so because the syntactic structure offers advice about what event the sentence describes, allowing the child to zoom in on relevant events in the environment. But identifying the precise nature of this advice in the lab requires a clear idea of the event concepts under which particular experimental scenes are represented. Prior work did not provide independent evidence for the conceptual representations under which children viewed their stimulus videos, which gave researchers an extra degree of freedom in explaining why children did not behave in accord with the One-to-One Matching hypothesis. When the predictions of the hypothesis were not borne out (in intransitive clauses), researchers had the freedom to say that infants had represented the events differently from (or more flexibly than) what the experimenters had intended. Here, we took that freedom away, giving empirical teeth to the experimental paradigm. In Experiment 1, we determined what concept the stimulus event in Experiment 2 was viewed under and what number of participants that concept was represented with. With better evidence for this representation, we were then in a better position to exploit a potential mismatch between argument number and participant number in testing our hypothesis.

Our results demonstrated that 20-month-olds allow a transitive clause to label an event that they represent as having three participants. This finding tells against views of syntactic bootstrapping based

in matching the number of syntactic arguments with the number of event participants. If bootstrapping inferences were driven by one-to-one matching of arguments and participants, then we would have expected our participants to take the two-argument sentence to label a two-participant concept, contrary to what we observed.

This is a welcome result, as Thematic Linking seems to us a more robust heuristic than One-to-One Matching. The mapping to thematic relations from the subject/object opposition is quite stable, both within a language and across languages, especially when conditioned on whether the clause describes an action, a change, or a mental state (Baker, 1997; Dowty, 1991). Thematic Linking therefore provides a sound and stable basis for inferring verb meanings. In contrast, the number of arguments does not always seem to match the number of participants, especially when the evidence for arguments excludes the meaning of the clause. Obvious examples of this, like passive in English, involve morphology that marks the mismatch. But these do not mark the limit of the challenge. It is not unusual for a language to have morphologically unmarked intransitives whose predicate expresses a 2-participant event concept; and some 3-participant event concepts find a happy home, we suggest, in ordinary transitive clauses.[5] One-to-One Matching is accordingly fallible: it will lead learners to wrong conclusions, in a systematic way. Thematic Linking, however, is more robust to variation in whether a participant is expressed as an argument, and more consistent with the facts of adult languages. Consequently, it allows for a continuous model of development, one that does not have to change the basis for inferring verb meanings as a specific language is acquired. We regard this as a significant advantage.

We also note that the argument we put forward here is not new to the syntactic bootstrapping literature. Content-based inferences between syntax and semantics are widely viewed as ubiquitous in word learning. Children infer that words used as nouns will name object kinds (Waxman, 1999; Waxman & Booth, 2001; Waxman & Markow, 1998), words used as verbs will name event kinds (He & Lidz, 2017), words used as determiners will have quantity-based meanings (Syrett et al., 2012; Wellwood et al., 2016), and words used as adjectives will have property-based meanings (Syrett et al., 2012; Waxman & Booth, 2001; Wellwood et al., 2016). The argument that learners might expect content-based correspondences between the syntax of a clause and their view of the event it describes is thus not particularly new or controversial. Our novel contribution is showing empirically that these inferences are available to children at the onset of verb learning. Specifically, we provide empirical evidence against the claim that young children's default bootstrapping strategy is based not on thematic content, but on one-to-one correspondence between syntactic and conceptual structure. This provides support for the view that richer content-based links between syntax and semantics underwrite children's earliest bootstrapping inferences, not merely in grammatical category learning, but also in acquiring verb meanings.

Of course, the advantages of Thematic Linking also come with a cost. Because these inferences require the learner to represent grammatical relations like "subject" and "object," or at least some syntactic asymmetry that underlies these informal rubrics, syntactic bootstrapping inferences must wait developmentally until the child has a basis for identifying these relations in the particular language they are acquiring. Word order and case morphology may serve as markers for these grammatical relations, but these markers differ across languages and therefore must be learned. In this respect, we are hopeful that learners might gain a foothold from indirect, albeit noisy, evidence available in the surface forms of sentences. One potential mechanism is prosodic bootstrapping: the syntactic divide between subject and predicate often correlates with a prosodic break cross-linguistically, an imperfect but potentially reliable source of information for learners (Christophe et al., 2008; de Carvalho et al., 2019; Morgan, 1986; Morgan & Demuth, 1996; inter alia). Another potential mechanism may make use of distributional asymmetries in the relative order of noun phrases and verbs, together with their

[5]While the current paper focuses on TAKINGS as a salient example, other 3-participant concepts – such as GIVINGS – may be less amenable to transitive descriptions (see again Tatone et al., 2015). Why these asymmetries arise, and how learners would treat the mapping between a transitive clause and a GIVING event in our experiment, are interesting questions that we leave for future work.

morphology. A learner who expects canonical clauses to have subjects, and expects that subject and object agreement will be marked in different ways, may be able to use the distributions and morphological marking of noun phrases in intransitive clauses as evidence for the position of subjects vs. objects in the language (Maitra & Perkins, 2023; Perkins & Hunter, 2023). Further mechanisms may make use of acquired noun knowledge: if children expect that animacy is a loose correlate of the subject-object asymmetry in transitive clauses, then identifying the referents of noun phrases in some transitive sentences may also help them learn how this grammatical asymmetry is marked (Becker, 2014; Childers & Echols, 2004; Lieven & Noble, 2011; Pinker, 1984; Scott & Fisher, 2009; inter alia). As Gleitman and colleagues have emphasized (Gillette et al., 1999; Snedeker & Gleitman, 2004), in order to use argument structure as a cue to verb meaning, the learner must first identify the phrases that function as arguments, even if only approximately. By the same token, we argue that the learner must be able to identify the subject and object of a clause, even if imperfectly, in order to engage the syntactic bootstrapper. How precisely this happens is an important open question, but our data suggests that by 20-months, children can do so.

In concluding, we would like to remark on the intuitive grip of One-to-One Matching. There has long been a view that the gross structure of sentences matches the gross structure of thoughts they express. For example, the grammatical division between subject and predicate is said to match a division in thought between a "logical subject" and a "logical predicate." One-to-One Matching is another instance of the genre. It suggests that, at least in the mind of the child, there is no more relational structure in the thought than there is in the sentence. This suggestion seems to resonate with textbooks from primary school to graduate school. Sometimes teachers claim that a sentence missing an argument is bad because it does not express a complete thought. Sometimes syntacticians require that syntactic features of a verb, its "theta roles," must be paired one-to-one with argument noun phrases in its domain. Sometimes semanticists insist that there is nothing for a phrase naming a predicate (or function) to do but combine with a phrase naming a thing of which it can predicate (or apply to). And sometimes logicians instruct us to transcribe "Edith sang the Marseillaise" as "S(e,m)."

But the resonances are deceptive, even if these sundry hypotheses were to be true. What is at issue for syntactic bootstrapping is not the combinatory regimes of verbs in the grammar, not even at the level of compositional semantics. It is, rather, how we conceptualize events. Even if the grammar were such that any transitive clause is centered on a verb with two syntactic "theta roles" that denotes a two-place function, this would have no necessary consequences for how we conceptualize those events that the sentence describes, except that it somehow involves two things. This would decide nothing about what further relations are entailed by the sentence (or its verb), much less which of these are psychologically distinguished as participant relations, in any independent sense of the term. The idea that argument relations match entailed relations in number has no hope of holding water, as it rests on a confound of content and structure, of entailment and structural proof. And the proposal that they match the psychologically distinguished participant relations, which is both possible and interesting, demands for its assessment a demonstration of which relations are salient in a learner's view of a stimulus event, independently of language. This is what we have endeavored to do. And what we have found is that young learners allow a clause to describe an event that they view as involving more participants than the clause has evident arguments. That is, we find that young children do not expect sentences to match thoughts in their structure. This suggests that more complex and nuanced correspondences between linguistic and conceptual representations underlie the early stages of language development.

Acknowledgments

We thank Rachel Dudley, Valentine Hacquard, Angela He, Mina Hirzel, Paul Pietroski, and Alexis Wellwood for valuable feedback and discussion, and we thank Gissell Alvarado, Abigail Kameny, Minqi Liu, and Tara Mease for their assistance in conducting these experiments.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Science Foundation: [#BCS-1551629], Doctoral Dissertation Improvement grant [#BCS-1827709], and NRT award [#DGE-1449815].

ORCID

Laurel Perkins http://orcid.org/0000-0003-0719-9510
Tyler Knowlton http://orcid.org/0000-0002-8207-0656
Alexander Williams http://orcid.org/0000-0002-7170-9143
Jeffrey Lidz http://orcid.org/0000-0001-8829-1495

Data availability statement

The data that support the findings of this study are available from the corresponding author, L.P., upon reasonable request.

References

Arunachalam, S., Escovar, E., Hansen, M. A., & Waxman, S. R. (2013). Out of sight, but not out of mind: 21-month-olds use syntactic information to learn verbs even in the absence of a corresponding event. *Language and Cognitive Processes*, *28*(4), 417–425. https://doi.org/10.1080/01690965.2011.641744

Arunachalam, S., Syrett, K., & Chen, Y. (2016). Lexical disambiguation in verb learning: Evidence from the conjoined-subject intransitive frame in English and mandarin Chinese. *Frontiers in Psychology*, *7*, 138. https://doi.org/10.3389/fpsyg.2016.00138

Arunachalam, S., & Waxman, S. R. (2010). Meaning from syntax: Evidence from 2-year-olds. *Cognition*, *114*(3), 442–446. https://doi.org/10.1016/j.cognition.2009.10.015

Arunachalam, S., & Waxman, S. R. (2011). Grammatical form and semantic context in verb learning. *Language Learning and Development*, *7*(3), 169–184. https://doi.org/10.1080/15475441.2011.573760

Arunachalam, S., & Waxman, S. R. (2015). Let's see a boy and a balloon: Argument labels and syntactic frame in verb learning. *Language Acquisition*, *22*(2), 117–131. https://doi.org/10.1080/10489223.2014.928300

Baker, M. C. (1988). *Incorporation: A theory of grammatical function changing*. University of Chicago Press.

Baker, M. C. (1997). Thematic roles and syntactic structure. In L. Haegeman (Ed.), *Elements of grammar* (pp. 73–137). Springer.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Becker, M. (2014). *The acquisition of syntactic structure: Animacy and thematic alignment* (Vol. 141). Cambridge University Press.

Brandone, A., Addy, D. A., Pulverman, R., Golinkoff, R. M., & Hirsh-Pasek, K. (2006). One-for-one and two-for-two: Anticipating parallel structure between events and language. *Proceedings of the 30th Annual Boston University Conference on Language Development*, Boston, MA (pp. 36–47).

Bunger, A., & Lidz, J. (2004). Syntactic bootstrapping and the internal structure of causative events. *Proceedings of the 28th Annual Boston University Conference on Language Development* ,Boston, MA (p. 74).

Bunger, A., & Lidz, J. (2008). Thematic relations as a cue to verb class: 2-year-olds distinguish unaccusatives from unergatives. *University of Pennsylvania Working Papers in Linguistics*, *14*(1), 4.

Carter, R. (1988). In B. Levin & C. Tenny (Eds.), *On linking: Papers by Richard Carter*. MIT Center for Cognitive Science Lexicon Project.

Childers, J. B., & Echols, C. H. (2004). 2½-year-old children use animacy and syntax to learn a new noun. *Infancy*, *5*(1), 109–125. https://doi.org/10.1207/s15327078in0501_5

Christophe, A., Millotte, S., Bernal, S., & Lidz, J. (2008). Bootstrapping lexical and syntactic acquisition. *Language and Speech*, *51*(1–2), 61–75. https://doi.org/10.1177/00238309080510010501

Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2004). *Habit X: A new program for obtaining and organizing data in infant perception and cognition studies (Version 1.0)*. University of Texas.

Csibra, G., Bíró, S., Koós, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science*, *27*(1), 111–133. https://doi.org/10.1207/s15516709cog2701_4
Davis, H. (2010). *Teaching grammar of St'át'imcets*. Unpublished ms.
Davis, H., & Demirdache, H. (2000). On lexical meanings: Evidence from Salish. In C. Tenny & J. Pustejovsky (Eds.), *Events as grammatical objects* (pp. 97–142). CSLI.
de Carvalho, A., Crimon, C., Barrault, A., Trueswell, J., & Christophe, A. (2021). Look! It is not a bamoule!: 18-and 24-month-olds can use negative sentences to constrain their interpretation of novel word meanings. *Developmental Science*, *24*(4), e13085. https://doi.org/10.1111/desc.13085
de Carvalho, A., He, A. X., Lidz, J., & Christophe, A. (2019). Prosody and function words cue the acquisition of word meanings in 18-month-old infants. *Psychological Science*, *30*(3), 319–332. https://doi.org/10.1177/0956797618814131
Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science*, *298*(5600), 2013–2015. https://doi.org/10.1126/science.1077066
Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, *67*(3), 547–619. https://doi.org/10.1353/lan.1991.0021
Fenson, L., Bates, E., Dale, P. S., Hartung, J. P., Pethick, S. J., Reilly, J. S., Reznick, J. S., & Thal, D. J. (1993). *The MacArthur communicative development inventories*. Singular Publishing Group.
Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In I. A. Sekerina, E. M. Fernández & H. Clahsen (Eds.), *Developmental psycholinguistics: On-line methods in children's language processing* (Vol. 44, pp. 97–135). John Benjamins Publishing Company. https://doi.org/10.1075/lald.44.06fer
Fillmore, C. J. (1968). The case for case. In E. Bach & R. Harms (Eds.), *Universals in linguistic theory* (pp. 1–25). Holt, Rinehart, & Winston.
Fillmore, C. J. (1970). The grammar of "hitting" and "breaking. In R. A. Jacobs & P. S. Rosenbaum (Eds.), *Readings in english transformational grammar* (pp. 120–133). Ginn and Company.
Fisher, C. (1996). Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive Psychology*, *31*(1), 41–81. https://doi.org/10.1006/cogp.1996.0012
Fisher, C., Gleitman, H., & Gleitman, L. R. (1991). On the semantic content of subcategorization frames. *Cognitive Psychology*, *23*(3), 331–392. https://doi.org/10.1016/0010-0285(91)90013-E
Fisher, C., Jin, K.-S., & Scott, R. M. (2019). The developmental origins of syntactic bootstrapping. *Topics in Cognitive Science*, *12*(1), 48–77. https://doi.org/10.1111/tops.12447
Fisher, C., & Song, H. (2006). Who's the subject? Sentence structures as analogs of verb meaning. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn the meanings of verbs* (pp. 392–426). Oxford University Press.
Foley, W. A., & Van Valin, R. D. (1984). *Functional syntax and universal grammar*. Cambridge University Press.
Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2016). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677–694. https://doi.org/10.1017/S0305000916000209
Gagliardi, A., Mease, T. M., & Lidz, J. (2016). Discontinuous development in the acquisition of filler-gap dependencies: Evidence from 15-and 20-month-olds. *Language Acquisition*, *23*(3), 1–27. https://doi.org/10.1080/10489223.2015.1115048
Gertner, Y., & Fisher, C. (2012). Predicted errors in children's early sentence comprehension. *Cognition*, *124*(1), 85–94. https://doi.org/10.1016/j.cognition.2012.03.010
Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science*, *17*(8), 684–691. https://doi.org/10.1111/j.1467-9280.2006.01767.x
Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135–176. https://doi.org/10.1016/S0010-0277(99)00036-0
Gleitman, L. R. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55. https://doi.org/10.1207/s15327817la0101_2
Gordon, P. (2003). The origin of argument structure in infant event representations. *Proceedings of the 26th Boston University Conference on Language Development*, Boston, MA (pp. 189–198).
Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General*, *142*(3), 880. https://doi.org/10.1037/a0030045
He, A. X. (2015). *Verb learning under guidance* [Doctoral dissertation]. University of Maryland College Park. https://d i.rg/drum.lib.umd.edu/handle/1903/16696
He, A. X., & Lidz, J. (2017). Verb learning in 14-and 18-month-old English-learning infants. *Language Learning and Development*, *13*(3), 335–356. https://doi.org/10.1080/15475441.2017.1285238
Hirsh-Pasek, K., & Golinkoff, R. M. (1996). The intermodal preferential looking paradigm: A window onto emerging language comprehension. In D. McDaniel, C. McKee & H. S. Cairns (Eds.), *Methods for assessing children's syntax* (pp. 105–124). The MIT Press.
Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*. MIT Press.
Jin, K.-S., Fisher, C., Orman, W., & Valleau, M. J. (2014). Early evidence for syntactic bootstrapping: 15-month-olds use sentence structure in verb learning. *Proceedings of the 38th Boston University Conference on Language Development*.
Kuhlmeier, V., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological Science*, *14*(5), 402–408. https://doi.org/10.1111/1467-9280.01454

Landau, B., & Gleitman, l. r. (1985). *language and experience: Evidence from the blind child*. Harvard University Press.
Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, *25*(3), 265–288. https://doi.org/10.1016/S0010-0277(87)80006-9
Levin, B., & Rappaport Hovav, M. (2005). *Argument realization*. Cambridge University Press.
Lidz, J., Gleitman, H., & Gleitman, L. (2003). Understanding how input matters: Verb learning and the footprint of universal grammar. *Cognition*, *87*(3), 151–178. https://doi.org/10.1016/S0010-0277(02)00230-5
Lidz, J., & Gleitman, L. R. (2004). Argument structure and the child's contribution to language learning. *Trends in Cognitive Sciences*, *8*(4), 157–161. https://doi.org/10.1016/j.tics.2004.02.005
Lidz, J., White, A. S., & Baier, R. (2017). The role of incremental parsing in syntactically conditioned word learning. *Cognitive Psychology*, *97*, 62–78. https://doi.org/10.1016/j.cogpsych.2017.06.002
Lieven, E., & Noble, C. H. (2011). The acquisition of argument structure. *Zeitschrift für Anglistik und Amerikanistik*, *59*(4), 411–424. https://doi.org/10.1515/zaa-2011-0409
Maitra, S., & Perkins, L. (2023). Filtering input for learning constrained grammatical variability: The case of Spanish word order. *Proceedings of the Society for Computation in Linguistics*, *6*(1), 108–120.
Marno, H., Farroni, T., Vidal Dos Santos, Y., Ekramnia, M., Nespor, M., & Mehler, J. (2015). Can you see what I am talking about? Human speech triggers referential expectation in four-month-old infants. *Scientific Reports*, *5*(1), 13594. https://doi.org/10.1038/srep13594
Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143–178. https://doi.org/10.1016/0010-0277(88)90035-2
Messenger, K., Yuan, S., & Fisher, C. (2015). Learning verb syntax via listening: New evidence from 22-month-olds. *Language Learning and Development*, *11*(4), 356–368. https://doi.org/10.1080/15475441.2014.978331
Morgan, J. L. (1986). *From simple input to complex grammar*. The MIT Press.
Morgan, J. L., & Demuth, K. (Eds.). (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Lawrence Erlbaum Associates, Inc.
Muentener, P., & Carey, S. (2010). Infants' causal representations of state change events. *Cognitive Psychology*, *61*(2), 63–86. https://doi.org/10.1016/j.cogpsych.2010.02.001
Naigles, L. R. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, *17*(2), 357–374. https://doi.org/10.1017/S0305000900013817
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*(5), 963–982. https://doi.org/10.1111/j.1551-6709.2011.01175.x
Papeo, L., Stein, T., & Soto-Faraco, S. (2017). The two-body inversion effect. *Psychological Science*, *28*(3), 369–379. https://doi.org/10.1177/0956797616685769
Pearl, L., & Sprouse, J. (2019). Comparing solutions to the linking problem using an integrated quantitative framework of language acquisition. *Language*, *95*(4). https://doi.org/10.1353/lan.2019.0084
Peña, M., Maki, A., Kovacic, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., & Mehler, J. (2003). Sounds and silence: An optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences*, *100*(20), 11702–11705. https://doi.org/10.1073/pnas.1934290100
Perkins, L., Feldman, N. H., & Lidz, J. (2022). The power of ignoring: Filtering input for argument structure acquisition. *Cognitive Science*, *46*(1). https://doi.org/10.1111/cogs.13080
Perkins, L., & Hunter, T. (2023). Noise-tolerant learning as selection among deterministic grammatical hypotheses. *Proceedings of the Society for Computation in Linguistics*, *6*(1), 186–198.
Perkins, L., & Lidz, J. (2020). Filler-gap dependency comprehension at 15 months: The role of vocabulary. *Language Acquisition*, *27*(1), 98–115. https://doi.org/10.1080/10489223.2019.1659274
Perkins, L., & Lidz, J. (2021). 18-month-old infants represent non-local syntactic dependencies. *Proceedings of the National Academy of Sciences*, *118*(41), e2026469118. https://doi.org/10.1073/pnas.2026469118
Piccin, T. B., & Waxman, S. R. (2007). Why nouns trump verbs in word learning: New evidence from children and adults in the human simulation paradigm. *Language Learning and Development*, *3*(4), 295–323. https://doi.org/10.1080/15475440701377535
Pinker, S. (1984). *Language learnability and language development*. Harvard University Press.
Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. MIT Press.
Pinker, S. (1994). How could a child use verb syntax to learn verb semantics? *Lingua*, *92*, 377–410. https://doi.org/10.1016/0024-3841(94)90347-6
Pozzan, L., Gleitman, L. R., & Trueswell, J. C. (2015). Semantic ambiguity and syntactic bootstrapping: The case of conjoined-subject intransitive sentences. *Language Learning and Development*, *12*(1), 14–41. https://doi.org/10.1080/15475441.2015.1018420
Rissman, L., & Majid, A. (2019). Thematic roles: Core knowledge or linguistic construct? *Psychonomic Bulletin & Review*, *26*(6), 1850–1869. https://doi.org/10.3758/s13423-019-01634-5
Saxe, R., Tenenbaum, J. B., & Carey, S. (2005). Secret agents: Inferences about hidden causes by 10-and 12-month-old infants. *Psychological Science*, *16*(12), 995–1001. https://doi.org/10.1111/j.1467-9280.2005.01649.x

Scott, R. M., & Fisher, C. (2009). Two-year-olds use distributional cues to interpret transitivity-alternating verbs. *Language and Cognitive Processes*, *24*(6), 777–803. https://doi.org/10.1080/01690960802573236
Scott, R. M., & Fisher, C. (2012). 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, *122*(2), 163–180. https://doi.org/10.1016/j.cognition.2011.10.010
Snedeker, J., & Gleitman, L. (2004). Why it is hard to label our concepts. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a Lexicon* (pp. 257–293). MIT Press.
Suttles, W. P. (2004). *Musqueam reference grammar* (Vol. 2). UBC Press.
Syrett, K., Musolino, J., & Gelman, R. (2012). How can syntax support number word acquisition? *Language Learning and Development*, *8*(2), 146–176. https://doi.org/10.1080/15475441.2011.583900
Tatone, D., & Csibra, G. (2020). Infants infer different types of social relations from giving and taking actions. *Proceedings of the Annual Meeting of the Cognitive Science Society*, Virtual.
Tatone, D., Geraci, A., & Csibra, G. (2015). Giving and taking: Representational building blocks of active resource-transfer events in human infants. *Cognition*, *137*, 47–62. https://doi.org/10.1016/j.cognition.2014.12.007
Tatone, D., Hernik, M., & Csibra, G. (2021). Facilitation of object encoding in infants by the observation of giving. *Scientific Reports*, *11*(1), 1–9. https://doi.org/10.1038/s41598-021-97910-3
Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science*, *7*(3), 270–276. https://doi.org/10.1111/j.1467-7687.2004.00345.x
Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*(2), 159–164. https://doi.org/10.1111/j.1467-7687.2007.00549.x
Waxman, S. R. (1999). Specifying the scope of 13-month-olds' expectations for novel words. *Cognition*, *70*(3), B35–B50. https://doi.org/10.1016/S0010-0277(99)00017-7
Waxman, S. R., & Booth, A. E. (2001). Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology*, *43*(3), 217–242. https://doi.org/10.1006/cogp.2001.0764
Waxman, S. R., Lidz, J. L., Braun, I. E., & Lavin, T. (2009). Twenty four-month-old infants' interpretations of novel verbs and nouns in dynamic scenes. *Cognitive Psychology*, *59*(1), 67–95. https://doi.org/10.1016/j.cogpsych.2009.02.001
Waxman, S. R., & Markow, D. B. (1998). Object properties and object kind: Twenty-one-month-old infants' extension of novel adjectives. *Child Development*, *69*(5), 1313–1329. https://doi.org/10.1111/j.1467-8624.1998.tb06214.x
Wechsler, S. (1995). *The semantic basis of argument structure*. CSLI.
Wellwood, A., Gagliardi, A., & Lidz, J. (2016). Syntactic and lexical inference in the acquisition of novel superlatives. *Language Learning and Development*, *12*(3), 262–279. https://doi.org/10.1080/15475441.2015.1052878
Wellwood, A., He, A. X., Lidz, J., & Williams, A. (2015). Participant structure in event perception: Towards the acquisition of implicitly 3-place predicates. *University of Pennsylvania Working Papers in Linguistics*, *21*(1), 32.
Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word–object associations by 14-month-old infants. *Developmental Psychology*, *34*(6), 1289. https://doi.org/10.1037/0012-1649.34.6.1289
Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, *3*(1), 1–30. https://doi.org/10.1207/S15327078IN0301_1
White, A. S., Hacquard, V., & Lidz, J. (2018). The labeling problem in syntactic bootstrapping: Main clause syntax in the acquisition of propositional attitude verbs. In *Semantics in acquisition. Trends in Language Acquisition Research (TiLAR)* (pp. 198–220). John Benjamins Publishing Company.
Williams, A. (2015). *Arguments in syntax and semantics*. Cambridge University Press.
Wilson, F., Papafraou, A., Bunger, A., & Trueswell, J. (2011). Rapid extraction of event participants in caused motion events. *Proceedings of the Annual Meeting of the Cognitive Science Society*, Austin, TX.
Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1–34. https://doi.org/10.1016/S0010-0277(98)00058-4
Yin, J., & Csibra, G. (2015). Concept-based word learning in human infants. *Psychological Science*, *26*(8), 1316–1324. https://doi.org/10.1177/0956797615588753
Yin, J., Tatone, D., & Csibra, G. (2020). Giving, but not taking, actions are spontaneously represented as social interactions: Evidence from modulation of lower alpha oscillations. *Neuropsychologia*, *139*, 107363. https://doi.org/10.1016/j.neuropsychologia.2020.107363
Younger, B. A., & Cohen, L. B. (1985). How infants form categories. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 19, pp. 211–247). Elsevier.
Yuan, S., & Fisher, C. (2009). "Really? She blicked the baby?": Two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science*, *20*(5), 619–626. https://doi.org/10.1111/j.1467-9280.2009.02341.x
Yuan, S., Fisher, C., & Snedeker, J. (2012). Counting the nouns: Simple structural cues to verb meaning. *Child Development*, *83*(4), 1382–1399. https://doi.org/10.1111/j.1467-8624.2012.01783.x

Appendix

Table A1. Audio stimuli, experiment 2.

Character Familiarization	Training: *Shak*
Look, it's a girl!	Look, the girl is gonna shake the bottle! She just shook the bottle!
Do you see her? There she is!	Wow, she's gonna shake it again! She just shook it again!
Now look, it's a boy!	Ooh, she's gonna shake the bottle! She just shook the bottle!
Do you see him? There he is!	Hey, she's gonna shake the bottle again! She just shook it again!
Can you find the girl? Where is she?	Uh-oh, she's NOT gonna shake THAT. She didn't shake THAT.
Now, where's the boy? Can you find him?	Yay, she's gonna shake the bottle! She shook the bottle!
	Now look, they're different!
	Find the one where she's shaking the bottle.
	Where is she shaking the bottle?
Training: ***Open***	**Training:** ***Give***
Look, the box is gonna open! The box just opened!	Look, the girl is gonna give the toy to the boy! She just gave the toy to him!
Wow, it's gonna open again! It just opened again!	Wow, she's gonna give it to him again! She just gave it to him again!
Ooh, the box is gonna open! The box just opened!	Ooh, she's gonna give the toy to him! She just gave the toy to him!
Hey, the box is gonna open again! It just opened again!	Hey, she's gonna give the toy to him again! She just gave it to him again!
Uh-oh, THAT's not gonna open. THAT didn't open.	Uh-oh, she's not gonna give THAT to him. She didn't give THAT to him.
Yay, the box is gonna open! The box opened!	Yay, she's gonna give the toy to him! She gave the toy to him!
Now look, they're different!	Now look, they're different!
Find the one where the box is opening.	Find the one where she's giving the toy to him.
Where is the box opening?	Where is she giving the toy to him?
Experimental Trial:	**Experimental Trial**:
Control Condition (Intransitive)	**Experimental Condition (Transitive)**
Look, the truck is gonna pim! It just pimmed!	Look, the girl is gonna pim the truck! She just pimmed the truck!
Wow, it's gonna pim again! It just pimmed again!	Wow, she's gonna pim it again! She just pimmed it again!
Ooh, the truck is gonna pim! The truck just pimmed!	Ooh, she's gonna pim the truck! She just pimmed the truck!
Hey, the truck is gonna pim again! It just pimmed again!	Hey, she's gonna pim the truck again! She just pimmed it again!
Uh-oh, THAT's not gonna pim. THAT didn't pim.	Uh-oh, she's not gonna pim THAT. She didn't pim THAT.
Yay, the truck is gonna pim! It just pimmed!	Yay, she's gonna pim the truck! She pimmed the truck!
Now look, they're different!	Now look, they're different!
Find the one where the truck is pimming.	Find the one where she's pimming the truck.
Where is the truck pimming?	Where is she pimming the truck?

Table A2. Video stimuli for training trials, experiment 2. Training trials follow the same format as the experimental trial (Table 1).

Trial	Familiarization	Contrast	Test
Shake	Girl shakes bottle of juice	Girl spins toy rattle	Girl shakes bottle/ Girl taps lid of bottle
Open	Girl opens box	Girl lifts toy house	Girl opens box/ Girl tilts box on its side
Give	Girl gives stuffed owl to boy	Girl throws ball, boy watches	Girl gives stuffed owl to boy/Girl hugs boy, holding owl