Thematic content, not number matching, drives syntactic bootstrapping

Laurel Perkins[1], Tyler Knowlton[2], Alexander Williams[3,4], Jeffrey Lidz[3]

[1]Department of Linguistics, University of California - Los Angeles

[2]MindCORE, University of Pennsylvania

[3]Department of Linguistics, University of Maryland

[4]Department of Philosophy, University of Maryland

Address for correspondence:

Laurel Perkins

3125 Campbell Hall

Los Angeles, CA 90095

perkinsl@ucla.edu

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 (L. R. 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". As support for this proposal, we provide evidence, first, that infants view certain scenes under a concept with three participant relations and, second, that toddlers readily accept a transitive clause as a description of that scene. Mappings between three-participant concepts and two-argument clauses argue against bootstrapping theories underwritten by mappings between structural features of syntactic and semantic representations, but are consistent with mappings between their contents, from syntactic positions to semantic relations.

This finding reinforces prior doubts about the empirical support for the structure-matching account. In particular, 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. Instead, our findings are more consistent with bootstrapping based on grammatical and thematic content: 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

Bootstrapping theories propose that correlations between clause syntax and predicate meanings may drive early verb and grammar learning, if children are aware that they exist and have access to the information needed to use them. A child who has evidence about meaning might use that information to draw inferences about the syntax of a clause (semantic bootstrapping: e.g. Pinker

Grimshaw, 1981; Pinker, 1984, 1989), and a child who has evidence about the syntax of a clause might use that information as a basis for inference about the meaning of words in that clause (syntactic bootstrapping: e.g. Gleitman, 1990; Landau & Gleitman, 1985). The viability of any form of bootstrapping depends on the robustness of the particular syntax-meaning correspondence relations that these inferences rely on, and on a child's ability to represent the relevant evidence at early stages of learning. A child's bootstrapping inferences will be ineffective if the syntax-meaning correspondence relations that she expects do not generally hold in the language that she is acquiring. And a child who expects syntactic and conceptual representations to relate in a particular way will not be able to apply that knowledge if she cannot represent the relevant syntactic information in sentences she hears, or if she does not perceive the world under a relevant conceptual description.

The 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, one that is orthogonal to the debate over whether syntactic or conceptual evidence drives initial bootstrapping inferences. Here, we aim to diagnose the specific correspondence relations that must underwrite these inferences: the knowledge at the heart of the inference engine, which is independent of the direction in which that inference is 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. Our test case is the correspondence principles that infants use to relate

clause syntax and event representations in order to infer the meanings of new verbs in simple transitive clauses.

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 (Pinker, 1984, 1989; Williams, 2015). This is similar in kind to an expectation that toddlers acquiring English seem to have about nouns versus verbs (A. X. He, 2015; A. X. 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 theoretical motivations and challenges for each of these hypotheses, and show how both can account for the existing behavioral data on early verb learning. Empirically distinguishing these hypotheses requires performing two steps that the literature has so far left implicit. First, in order to ask how young children put their sentence and scene representations into correspondence for verb learning, we need to test the structures under

which they represent particular stimulus scenes, independent of language. Second, we need to test scene-sentence pairings for which One-to-One Matching and Thematic Linking would predict different inferences about the meanings of verbs in those sentences. We provide both of these tests in the current study.

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, and take the "participant relations" to be a privileged subset of the relations entailed by an event concept. Relations entailed by a concept are all those that must hold if the concept applies.[1] The concept of CYLINDER entails the relations of DIAMETER and HEIGHT, for example, since anything to which CYLINDER applies must enter the latter two relations to certain lengths.

[1] To be precise, entailment relates truth-bearing propositions or formulas, not concepts or relations: A entails B just in case the circumstances where A is true are subset of those where B is. So when we say that a concept entails a relation, we actually mean that, if the concept applies to a certain *x*, this entails that something enters that relation to *x*. For example, if we say that that a one-place event concept C entails a two-place relation R, we mean that, for any e, C(e) entails $\exists x[R(e,x)]$.

But we cannot use "participant relation" as a synonym for "entailed relation." If we did, 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 *f* that materially instantiates what was sung, and so on. So if the notion of a participant relation is to do any work, it must distinguish just a few relations from among all of those that a concept entails. The participant relations, as such, must be somehow central to our conception of events that fall under it, or to their metaphysics. We will use the term to indicate a psychological distinction. The participant relations are those an event concept represents explicitly, or makes salient, when it is tokened.

For convenience we formalize this in terms of the valence (arity, adicity, degree) of the concept, which we will indicate with a sequence of variables within angle brackets.[2] The one-place concept SINGING<e>, for example, has no participant relations, while the two-place SINGING-BY<e,x> has one in the role of singer, and the three-place SINGING-BY-OF<e,x,y> adds one more in the role of sung. By assumption, all three concepts in this example share the same entailments, 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 represent explicitly, and therefore, given our notation, in which they represent as participant relations. For convenience we will refer to multi-place event concepts by way of an analysis that labels each of

[2] The angle brackets in our notation can be understood as indicating that each variable in the sequence is bound by a lambda in an extensionally equivalent formula of the lambda calculus, so that the 2-place concept SINGING-BY<e,x> corresponds in extension to the lambda term λxλe[True if and only if *e* is a singing by *x*], which names a function that maps a thing *x* and an event *e* to a truth value.

its participant relations with a conventional name for a general thematic relation, separating these names with hyphens. Instead of writing SINGING-BY-OF<e,x,y>, for example, we will write SINGING-AGENT-THEME<e,x,y>.[3] Whether this concept has exactly the internal structure suggested by our name for it (a composite of the SINGING, AGENT and THEME concepts) is something we cannot answer here. We require only that the participants roles correspond to 'slots' or 'variables' in the concept, and are explicitly represented in at least that sense.

Lastly, it is essential to separate a claim about what concept a verb "expresses," for purposes of bootstrapping, from one about the "semantic value" or "denotation" of that verb itself. When we say in this paper that a verb "expresses" a concept, we commit only to what matters to bootstrapping inferences: that the learner, in their learning experience with that verb, view the event described by its clause under that concept. We might also expect that the concept a verb "expresses" may be the representation that we generally have of events when we label them with that verb. But this is not essential to the acquisition question. It is likewise a further question whether the concept comes to be associated lexically with the verb in the acquired grammar, as an intrinsic property of the word. And in any case, the valence of a concept under which an event is viewed does not decide the compositional semantics of a clause used to describe it, or specifically the semantic value of its verb. There is no logical conflict between saying that a verb expresses, in our sense, a triadic event concept, V<e,x,y>, with two participant roles, and saying that its semantic value is simply a one-place predicate of events, true of some events and false of others. In that case, any participant relations that are associated with the subject or object would be introduced compositionally by parts of the structure other than the

[3] This corresponds in extension to the lambda term λyλxλe[True if and only if Singing(e) & Agent(e,x) & Theme(e,y)].

verb (Dowty, 1991; Kratzer, unpublished m.s.; Williams, 2021); and any that are not associated with an argument, if this turns out to be possible, are understood as existentially bound.

Bootstrapping from Syntactic Position

Bootstrapping hypotheses under the family of Thematic Linking propose that children exploit correspondences between grammatical and thematic relations, which relate in reliable ways across languages (e.g., Pinker, 1984, 1989). Here, we abstract away from the details of how these correspondences should be characterized, a topic with a substantial literature (Baker, 1988, 1997; Carter, 1976; Dowty, 1991; Fillmore, 1968, 1970; Foley & Van Valin, 1984; Jackendoff, 1972; Levin & Hovav, 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 verbs like *kick* to occur in basic clauses where the subject names the kicker and the object names the thing kicked: e.g., *Kim kicked the table*. Cross-linguistically, these kinds of verbs do not tend to distribute with the reverse linking pattern, in which the agent of the action is expressed by the object and the patient is expressed by the subject. For instance, verbs that mean ‘kick’ do not tend to occur in sentences like **The table kicked Kim*, in which *the table* expresses the thing kicked and *Kim* expresses the kicker. Generalization (1b) and (1c) capture cross-linguistic tendencies for agents vs. patients to be realized as clause arguments depending on the type of event that the clause describes. Verbs

that describe events of change tend 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'). Conversely, verbs that describe an action of an agent tend to distribute with arguments expressing that agent, though this tendency is much weaker than the first two (see the following section). 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 the hypothesis that children bring certain expectations about the alignments between syntactic position and thematic relations is its ability to account for these cross-linguistic generalizations (Baker, 1997; Dowty, 1991; Pinker, 1984, 1989). Suppose that children bring to the learning task some broad expectations about the links between particular grammatical and thematic relations in natural language grammars— for instance, knowledge of the principles 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 might help explain why similar patterns are exhibited in diverse languages: they arise from grammatical principles that form part of learners' initial linguistic capacities at the onset of language development, and thus shape the grammatical inferences that learners draw on the basis of the data that they are exposed to.

These linking principles could drive syntactic bootstrapping inferences in the following way. Children might be able to use evidence about the likely thematic content of clause arguments to infer that a clause describes an event with thematic relations borne by the

individuals that those arguments name. Suppose a child hears an unknown verb in a transitive sentence, e.g. *The girl VERBed the truck*, and she represents *the girl* as the subject and *the truck* as the object of the clause. She might then use knowledge of generalization (1a) to infer that this clause describes an event in which a girl is an agent and a truck is a patient. Provided that her nonlinguistic cognitive capacities yield event percepts in which participants are represented as more vs. less agentive, this might allow her to narrow down the candidate events to those of girls acting on trucks, and thus narrow down the range of possible meanings for the verb.

The viability of Thematic Linking for early bootstrapping rests on two assumptions about young learner's 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. We use 'subject' and 'object' as useful descriptive rubrics for the relevant asymmetry in grammatical relations; likewise, '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 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).

Diverse experimental findings suggest that this first assumption is borne out. Infants as young as 6-12 months represent an asymmetry between the participant relations that we describe in terms of 'agent' and 'patient': they represent agents as likely to be animate and able to initiate

motion (Leslie & Keeble, 1987; Saxe et al., 2005; Woodward, 1998), able to cause changes to other objects and individuals (Muentener & Carey, 2010), goal-directed (Csibra et al., 2003; Kuhlmeier et al., 2003; Wagner & Carey, 2005; Woodward, 1998), and rational (Csibra et al., 2003; Gergely et al., 2002; Meltzoff, 1988). Another body of research finds that adults are able to perceive asymmetries among agents, patients, and other event participants on the basis of extremely brief visual exposure to events (under 100ms), indicating that participant relations may be encoded at very early stages of human visual processing (Hafri et al., 2013; Wilson et al., 2011). Thus, it is likely that young learners' nonlinguistic conceptual representations are rich enough to support bootstrapping to and from specific participant relations.

There is much less evidence, however, bearing on the richness of infants' syntactic representations at early 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 in the speech stream: languages mark the syntactic relations that underlie these rubrics in highly variable ways. So if the generalizations are to be explained by linking principles that are used to bootstrap verb meanings, learners would need evidence, in perceived features of the sentences that they hear, for the syntactic relations that we describe under the rubrics of subject and object. 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 by using rougher and less accurate 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.[4]

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 of the clause. However, other mechanisms for identifying subjects and objects in clauses may be available. For instance, some work has argued that 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 in intransitive sentences, may serve as the basis for distributionally-based inferences about the canonical position of subjects vs. objects (Perkins & Hunter, in prep). 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. Their ability to do so will determine when a strategy like Thematic Linking could be used, and whether it is available to drive the initial stages of verb learning.

Bootstrapping from Argument Number

The theoretical challenges to Thematic Linking have motivated another influential bootstrapping proposal. If infants at the onset of verb learning cannot yet differentiate the grammatical relations

[4] Because we use "subject" and "object" as informal descriptive rubrics, when we say that a learner may use linking principles stated in these terms, we require only that they access some high-level syntactic classifications consistent with these rubrics, whatever the specific syntactic relations are that actually exemplify them in the child or adult grammar.

of clause arguments, then they would need a way to relate coarser properties of their linguistic representations to properties of their event representations. One useful property could be the number of clause arguments that they represent in a sentence. Suppose infants had a way to identify which expressions in a clause are arguments— for instance, they might guess that noun phrases are likely candidates. Then even this minimal information might be used to draw an inference about how the clause represents its event, if they moreover expect the clausal arguments to 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).[5] If 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 another, rougher 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 has gone by various names in the literature, including "Structure-Mapping" (Fisher et al., 2019) and "Participant-Argument Matching" (A. X. He, 2015; Williams, 2015).

[5] Some recent proposals have moved from bi-directional one-to-one participant-argument alignment to a unidirectional mapping, whereby each argument in a sentence is assumed to match a perceived event participant, but not necessarily vice-versa (Fisher et al., 2019). Below, we discuss the consequences for learning that would arise from adopting this weaker version of the number-based bootstrapping strategy.

For purposes of the current discussion, we call it "One-to-One Matching": 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 any more specific information about the syntactic position of arguments or the content of the participants' relation to their event. For example, suppose again that an infant hears a transitive sentence with an unknown verb— *The girl is VERBing the truck*— and can identify that the sentence has two noun phrase arguments, but cannot correctly represent one as the subject and the other as the object. One-to-One Matching would still provide her with some basis for inference about meaning: it would tell her that this sentence describes an event that she perceives as having two participants, not one or three. If she readily represents only some candidate events under 2-participant concepts, then she might be able to narrow down the range of events that the novel verb labels.

One-to-One Matching is an attractive hypothesis because it would seem to provide strong constraints on verb learning, and yet is quite simple to implement. If children expect noun phrases to correspond to participants, then they need only count the number of noun phrases in a clause in order to infer the participant structure of the event representation expressed by that clause.[6] A sentence with *n* noun-phrases-as-arguments will describe an event as having *n* participants. Children do not need to access any finer-grained syntactic or semantic information: in particular, they do not need to parse the sentence or infer the thematic relations of the

[6] While the literature commonly refers to this strategy as 'counting,' it may be more accurately understood as a process of identifying the NPs in a sentence and finding an event representation in which those NPs can be put into one-to-one correspondence. That is, infants do not actually need the ability to count; they only need the ability to hold clause arguments in memory in order to pair them with an event representation.

arguments. Thus, this strategy has the potential to be useful even for children with very immature knowledge of their language.

But the utility of this strategy in the absence of finer-grained syntactic or semantic information rests on several key assumptions. First, children must be able to stably perceive the number of arguments in sentences they hear. If they mis-identify the number of arguments in a clause, One-to-One Matching may lead to faulty inferences about meaning. Challenges for this hypothesis come from empirical evidence of occasionally inaccurate sentence representations early in development, where infants may either mis-identify noun phrases as core clause arguments when they in fact are not (Gertner & Fisher, 2012), or may fail to recognize clause arguments that are present in non-canonical word orders (Gagliardi et al., 2016; Perkins & Lidz, 2020, 2021). In order for One-to-One Matching to provide helpful guidance early in development, 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, Feldman, & Lidz, 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, and are, very importantly, stable in participant number. Plausibly, infants view the events around them through roughly the same conceptual lens as all human beings do (Baillargeon, 2004; L. Gleitman & Papafragou, 2005; L. R. Gleitman, 1990; Spelke & Kinzler, 2007), as supported by the literature on representations of participant relations discussed above. However, some experimental work suggests that infants might not always view stimulus events under the same representations intended by the experimenters (Brandone et al., 2006; Pozzan et al., 2015). Furthermore, if children viewing a particular event can flexibly 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, then a number-based strategy will provide little help in determining which event out of many a speaker's sentence describes (Wellwood et al., 2015; Williams, 2015).

For example, consider what might happen if a child could readily represent a particular 'pushing' scene either as a pushing with two participants, a relation between a pusher and the thing pushed, or as a pushing with only one participant, a predicate true of anything pushed. That is, suppose a child easily switches from representing this 'pushing' scene as PUSHING-BY-OF<e,x,y>, "a pushing by *x* of *y*," to representing it as PUSHING-OF<e,y>, "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 from a 2-participant to a 1-participant description of a particular event, while maintaining all of the same entailments of the event concept, then the One-to-One Matching strategy would offer much less helpful advice. A transitive clause might describe the 2-participant representation of this PUSHING event, or any other event that can be suitably down-shifted to a 2-participant representation. An intransitive clause might describe the 1-participant representation of this PUSHING event, or any other event that can be suitably down-shifted to a 1-participant representation. That is, the child would be unable to distinguish among any event types that could be represented as having more than one participant, because they could all be represented, in addition, as having just one. In order for a number-based strategy 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).

The second assumption is that children expect bi-directional matching: arguments in a clause must be in one-to-one correspondence with perceived event participants. Consider a weaker, uni-directional version of this hypothesis: 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). Here, a clause with *n* arguments should be able to label any event readily perceived with *n* or more participants. This strategy would only be constraining enough to guide verb learning under very specific circumstances (Williams, 2015). A learner who hears a new verb in a transitive clause could infer that it does not describe an event seen with only one participant. But a verb in an intransitive clause could describe an event seen with one participant, or two, or three: argument number would provide very little guidance in this case, and inferences about verb meaning would need to be drawn from another source of information, e.g. from the inferred thematic content of the argument. For a number-based strategy to provide strong enough constraints to guide verb learning in the general case, in the absence of any more fine-grained linguistic information, it must be formulated as an expectation of one-to-one correspondence between participants and arguments, for any number.

This assumption faces an immediate empirical challenge. It would seem to warrant incorrect inferences with some regularity, since there sometimes seem to be fewer arguments participants. For instance, 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.[7]

[7] This is one reason why we might expect that a verb in general expresses a concept with a specific number of participants, regardless of how many arguments occur in its clause, and regardless of its role in the compositional semantics.

Short passives, like "The truck was VERB-ed," 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.[8] 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. To avoid error, the One-to-One Mapper would have to allow that participant structure varies with argument structure, even holding the verb fixed. But then the mapping principle would be much less useful: it would predict nothing about the meaning of the verb for intransitive clauses, and for transitive clauses, only that the verb here expresses a concept with at least two participant roles. But that is to say no more than that argument NPs in general name event participants, a far weaker theory.

The challenge does not end once knowledge of the passive is acquired. Even in basic active clauses of English, with no morphological sign of an '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. Imagine a scene in which a girl takes a toy truck from a boy. It is plausible that this scene might be viewed under a 3-participant concept, one in which the girl, truck, and boy fill participant roles. But it also 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

[8] 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".

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 an expectation of one-to-one matching between participants and arguments would lead young learners astray in this case. Learners using this strategy would erroneously conclude that *The girl took the truck* cannot describe an event viewed as 3-participant TAKING, in which the victim/source participant is explicit in the representation. Instead, such a learner would be led to conclude that the sentence describes a 2-participant event, such as picking up or moving the truck, and thus acquire the wrong meaning for *take*. Importantly, the problem cannot be avoided by imagining that a learner would impute a silent third argument NP to the clause with *take*, since there is no morphosyntactic or pragmatic evidence for doing so. Learners or linguists who would posit a silent third argument could do so only on the basis of a hypothesis about the concept expressed by *take*, using One-to-One Mapping to bootstrap from semantics into syntax. But the syntactic bootstrapper cannot do that, since the meaning is exactly what they are trying to acquire.

This sort of problem becomes more salient once we broaden our survey of languages beyond English. In particular, many languages allow predicates that entail an agent, and in English would require transitive syntax, with meanings like “crush” or “repair”, to occur in simple intransitive clauses without an argument NP realizing the entailed agent, without anything like passive marking. Examples like this can be found, for example, in Mandarin, Igbo, Fijian and Hindi (Williams, 2015). An extreme case is St’át’imcets, a Salish language of western Canada. In St’át’imcets nearly every verb root can occur, in its bare form, in a simple intransitive clause (Davis, 2010; Davis & Demirdache, 2000). Consider the following example:

(2) Qámt kwskwimçxen

hit.with.projectile det.NAME

Approximately: ‘Kwimçxen got beaned.’ (Davis, 2010)

This entails that someone got hit with a thrown object, but the clause has neither the hitter or the thrown object as syntactic arguments. It is intransitive, and realizes only the role of 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, and likewise about any 2-or-more participant verb in this intransitive context, which is available to nearly every verb in the language. Thus the assumption of one-to-one correspondence would provide rather unhelpful advice for St’át’imcets learners in general (Williams, 2015).

Thus One-to-One Matching 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, prior to mastering alternations like passive. Consequently the strategy has generally been proposed only as an initial heuristic to get verb learning started, which would be abandoned as children acquire more sophisticated knowledge of their language (Fisher, 1996; Lidz et al., 2003; Lidz & Gleitman, 2004). But this means that the pathway to a mature grammar involves transiting through an errorful stage in which children’s early beliefs about the correspondence between

syntax and semantics does not resemble those of adults. Therefore, this hypothesis requires a further account of how children's grammatical knowledge develops past this stage.

Prior Empirical Findings

Previous tests of One-to-One Matching have mostly used preferential looking designs based on a seminal study by Naigles (1990). In this task, 25-month-olds were presented with a novel verb in the context of two scenes: one intended to be viewed as a causal event of a duck pushing over a bunny, and one intended to be viewed as a non-causal event of a duck and a bunny each wheeling their arms independently. Infants who heard the novel verb in a transitive clause (*The duck is gorping the bunny*) looked longer at the 'pushing' scene then infants who heard an intransitive clause (*The duck and the bunny are gorping*). 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. If they perceive a pushing scene but not an arm-wheeling scene under a two-participant concept, then, given these two choices, they should draw the inference that *gorping* most likely labels PUSHING rather than ARM-WHEELING. Further support for One-to-One Matching has come from a large body of work replicating this basic finding, in which infants as young as 19 months robustly prefer an intended 2-participant event when they hear a transitive clause (Arunachalam et al., 2013; Arunachalam & Waxman, 2010; Messenger et al., 2015; Yuan et al., 2012; Yuan & Fisher, 2009). 15-month-olds show this same preference when tested with simplified visual stimuli (Jin & Fisher, 2014).

A complication arises from children's behavior with intransitive clauses. Beyond Naigles' original (1990) study, 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). Because these results are not predicted under One-to-One Matching, several methodological explanations have been proposed. Infants may not perceive the presented sentence as intransitive. To control the number of nouns across conditions, several studies used intransitive sentences with conjoined subjects (e.g. *The duck and the bunny are gorping*), which some have argued may be mistaken for two separate arguments of the clause (Gertner & Fisher, 2012; Yuan et al., 2012). If so, One-to-One Matching would guide infants towards the event intended to be viewed with two participants (henceforth "2-participant event"). Alternatively, it is possible that infants do not reliably perceive the presented scenes with the intended number of participants. A scene intended to be viewed as one person pushing another might also be viewed under substantively 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.

However, argument number is not the only source of information that young infants appear to use in their early bootstrapping inferences. Infants as young as 17-21 months show sensitivity to the canonical word order of subjects and objects in their language and awareness of how different arguments, subject or object, link to different participant roles (Gertner et al., 2006; Hirsh-Pasek & Golinkoff, 1996; Lidz et al., 2017). For instance, in Gertner et al. (2006), 21-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, thus appearing to link the argument in subject position to the agent in an event and the argument in object position to the

patient.[9] In this task, a number-based strategy is not particularly helpful: this strategy predicts that a transitive clause describes a 2-participant event, but does not by itself predict which 2-participant event, or which roles the arguments link to. In order to succeed in this task, infants needed information not merely about the number of clause arguments, but about their likely grammatical and thematic relations.

Another line of work has shown that children can draw even finer-grained information about event type from the thematic content of intransitive subjects (Bunger & Lidz, 2004, 2008; Scott & Fisher, 2009). Recall that the subject of an intransitive clause can label either an agent (e.g. *Kim swept*) or a patient (e.g. *The vase broke*). These differences correlate with differences in event type, as characterized above in (1b-c): intransitives whose subject is capable of agency will likely describe actions in which that subject is the agent, whereas intransitives whose subject is incapable of agency will likely describe a change in which that subject is the patient (e.g., Fillmore, 1970; Levin & Hovav, 2005; Williams, 2015). Children as young as 2 years old show awareness of these patterns (Bunger & Lidz, 2004, 2008; Scott & Fisher, 2009). For instance, the 28-month-olds tested by Scott and Fisher (2009) inferred that a novel verb in an intransitive clause with an inanimate subject (e.g. *The pillow dacked*) was more likely to label an event in which a change is effected on a patient. Conversely, they inferred that a novel verb in an intransitive clause with an animate subject (e.g. *He dacked*) was more likely to label an action of an agent that does not effect a change. As above, argument number does not provide the information needed to perform this inference: children could only succeed on this task by using cues to the thematic relation of the intransitive subject, such as animacy. It appears that when the intransitive subject was animate, they inferred that it was a likely agent and the sentence

[9] Note that these results do not tell us how infants differentiate these grammatical relations: whether by using particular structural positions in a clause, or by using their linear order.

described an action of that agent; when the subject was inanimate, they inferred that it was a likely patient and the sentence described a change undergone by that patient.[10]

If infants can exploit generalizations between syntactic positions and thematic relations when the number of arguments in a sentence is uninformative, one might wonder what role, if any, argument number plays in their bootstrapping inferences in other contexts. A bootstrapping strategy relying on thematic content would appear to be equally compatible with the findings of the prior preferential looking literature. If infants are able to identify which argument in a transitive clause is the subject and which is the object, and they moreover expect subjects to name agents and objects to name patients, then they will prefer an event readily viewed under a 2-participant description, in which an agent acts on a patient, over an event readily viewed as only having one participant. And when infants are indifferent between these two event types when hearing a novel 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. Animacy has not been a useful cue in most prior tests of One-to-One Matching, which have typically deployed animate actors in all participant roles. If infants in these studies were unable to determine the likely thematic relation of an intransitive subject, the sentence would be ambiguous in context, predicting no looking preferences at test.

Thus, prior experimental work leaves open many possibilities regarding the role of number vs. thematic content in early bootstrapping. Infants' sensitivities to clause transitivity in verb learning tasks might be explained without referencing argument number, appealing instead to more specific correspondences between grammatical and thematic relations. It is possible that

[10] Note that animacy is only loosely correlated with thematic relations. Unaccusative verbs can take both animate and inanimate subjects, and likewise for unergative verbs. Nonetheless, children appear to use the animacy of noun phrases as a probabilistic cue to infer thematic and syntactic structure in a variety of experimental tasks (Becker, 2015; Becker et al., 2012; Bunger & Lidz, 2004, 2008).

both a number-based and a thematic content-based strategy could co-exist in early development, and learners make use of either when convenient (Fisher et al., 2019). It is also possible that children primarily rely on a number-based heuristic at early stages of syntactic development, but come to deploy more sophisticated bootstrapping strategies once they have acquired richer grammatical knowledge (Fisher, 1996; Lidz & Gleitman, 2004). To differentiate these alternatives, we need to know whether infants initially privilege the numbers of arguments and participants in their bootstrapping inferences.

Because One-to-One Matching and Thematic Linking can account for the same behavior in prior tasks, our test case must be one in which these strategies would lead to different inferences about verb meanings. The crucial scenario is one in which infants hear a sentence with fewer arguments than participants they represent in the accompanying scene. A number-based strategy predicts that they should register a mismatch: this sentence should not be a good fit for the scene as they initially represented it. A thematic-content based strategy 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. But 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.

The current work differentiates these strategies in the following way: 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 the One-to-One Matching hypothesis that has not yet been previously examined: this strategy should hold for any number of arguments and participants, including those greater than two. We first 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. We then ask whether 20-month-olds allow that 3-participant concept to be labeled by a 2-argument sentence with a novel verb: *The girl pimmed the truck*.

One-to-One Matching predicts that infants should think this sentence cannot be a description of this scene, as they readily perceive it. The sentence has only two arguments, and thus must describe a nearby event concept that they represent as having only two participants—perhaps the girl's grabbing or obtaining of the truck, independent of the boy. But Thematic Linking predicts no mismatch. If infants represent *the girl* as subject and *the truck* as object, then they should think that this sentence describes an event in which the girl is agent and the truck is patient. This means that the sentence could be a possible description of the 3-participant TAKING under which they readily view this scene, provided that they perceive those particular participant roles; it does not matter that a third perceived participant, the boy, is not mentioned. Examining infants' conclusions about verb meaning in this case study allows us to tell apart number-driven bootstrapping from more flexible inferences based on thematic content. But before we can test the relative merits of these bootstrapping strategies, we must first determine the conceptual representations under which infants perceive the events used in our task.

Diagnosing Event Representations

Identifying infants' bootstrapping strategies is only possible if we first fix the conceptual representations under which infants perceive events in a particular experimental task. If we do

not know the number of participants that infants readily perceive in a particular scene, then we cannot tell whether that conceptual representation matches or mismatches the number of arguments that they hear in a particular sentence. The evidence for these representations cannot come from the verbs and argument structures that adults would use to describe a particular event, because this is the knowledge that children are trying to acquire; we need a diagnostic that is independent of language.

Any particular event concept entails many relations involving that event. For instance, 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. Any taking has all of these aspects, just like any cylinder has a height and a diameter. But not every relation entailed by a predicate corresponds to an argument of that predicate. The same could be said about our psychological representations of predicates. Not all entailed relations will be on par with each other psychologically, because any particular psychological representation will foreground some of those relations and background others. Marr (1982) provides a useful analogy: we can write down the numerical content THIRTY-SEVEN as the base-10 '37' or as the base-2 '100101'. Whereas the former representation makes decomposition into powers of 10 explicit, the latter representation backgrounds this information but makes decomposition into powers of 2 explicit. In the same way, not all of the relations that are entailed when we describe something as a TAKING will be psychologically privileged to the same extent. Only some will be explicit in the conceptual structure under which we represent that taking event: these are what we call the "participant relations," as stipulated above (Williams, 2015). Our first question is, for any given stimulus scene, which relations are privileged in the concept that infants view it under? What are its participant roles?

There are many possible ways of writing down the structure under which infants could view a particular taking event. Here are just some of these possibilities, in which each participant role is given a separate label for ease of reference:

(3) a. TAKING-AGENT-PATIENT-SOURCE<e,x,y,z> 3-participant

b. TAKING-AGENT-PATIENT<e,x,y> 2-participant

c. TAKING-AGENT<e,x> 1-participant

Infants might view a particular taking event under a 3-participant conceptual 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 possibilities (Tatone et al., 2015, 2021). These different possibilities for the structure of an event representation exist independently of the entailments of the event concept. For instance, infants may represent a scene under a concept that entails a source, without necessarily representing the source as a participant that is explicit in their conceptual structure. Put another way, each of the representations in (3) are true of all and only the same events. They differ only in which participants are explicitly encoded in the representation: (3b-c) do not make the source explicit as a participant. So in order to diagnose the number of participants that infants represent (and use in syntactic bootstrapping), we need to determine which relations they view not merely as entailments of the predicate, but as psychologically privileged entailments.

Recent work has developed a new method for diagnosing the structure of infants' event representations, building off of a task introduced by Gordon (2003). In this study, 10-month-old infants were habituated to a silent scene that could be plausibly seen as a GIVING event: a girl gave a toy to a boy. At test, infants either saw another token of the same scene, or they saw a

scene in which the motion of the two actors was constant but the toy was no longer present: the girl approached the boy with empty hands. In a control condition, a separate group of infants were habituated to a silent scene in which the girl hugged the boy while holding the toy. At test, the same manipulation was performed: infants either saw another token of the 'hugging-with-toy' scene, or they saw the girl hug the boy in the same way without holding the toy. Results showed that infants dishabituated (recovered attention) to the disappearance of the toy from the giving scene, but not from the hugging scene. These results may be explained in the following way. If infants represented the 'giving' scene under a concept that entails a gift, then the change at test is a large conceptual change: there is nothing filling that entailed relation. But if infants represented the 'hugging' scene under a concept in which the toy does not fill an entailed relation, but is merely incidental, then the change at test is smaller conceptually: the hugging-without-toy scene can still be described as a HUGGING with all of its entailed roles present. Because the same physical manipulation— removing a toy— led to different patterns of dishabituation for the two types of events, we might infer that infants considered the toy more central to one of these events than to the other. All else being equal, this difference potentially arose because infants represented these scenes under concepts with different entailments.

Thus, in a controlled design, infants' patterns of attention to a subtle change in a stimulus event can provide a window into the conceptual representation under which they view that event. However, the result from Gordon (2003) doesn't tell us whether infants were responding merely to a change in conceptual entailments, or moreover to a change in participant structure. Infants appeared to view the 'giving' scene under a concept in which the toy filled an entailed relation. But this relation may or may not have been represented as a participant, explicit in the structure of their representation. Infants may have viewed this scene under a 3-participant concept in

which the toy filled a psychologically privileged relation, or they may have viewed it under an otherwise equivalent 2-participant concept that entails something given, without that relation being psychologically privileged.

To tease apart these possibilities, we use a method first introduced in Wellwood et al. (2015) and He (2015). The logic of this method is as follows: if infants seem to notice changes to the hypothesized participant structure under which an event is viewed, above and beyond changes to the physical properties of the event, then we might take this as evidence about the structure under which they had viewed the event. He (2015) habituated infants to a silent scene in which a girl jimmies open a box using a lever. At test, infants saw one of two subtle changes. In one condition, the girl now opened the box with her hand; the lever was still present and visible, but no longer used as an instrument of opening. In a second condition, girl continued 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 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. Infants' differential response to these two types of changes 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 in their event representation.

Here, we extend this paradigm to diagnose the participant structure under which infants view a 'taking' stimulus scene: a girl takes a truck from a boy. We first show that infants readily represent this scene under a concept that has three participants (Experiment 1). We then show

that young verb learners allow this 3-participant event concept to be labelled by a clause with only two arguments (Experiment 2). This behavior is consistent with Thematic Linking but inconsistent with One-to-One Matching. Thus, this sequence of experiments empirically distinguishes a number-matching bootstrapping strategy from its alternatives, suggesting that infants privilege the grammatical and thematic relations of arguments above argument number in their early verb learning inferences.

Experiment 1

Our first experiment adopted the habituation-based task introduced by Wellwood et al. (2015) and He (2015) to test 10-month-olds' nonlinguistic representations of a 'taking' scene. We chose this age range 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. We habituated infants to a silent video in which a girl picks up a toy truck and moves it towards 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. It is plausible that infants would 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 towards 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.

Method

Participants

Participants included 32 typically-developing infants (16 males, 16 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.

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 Samsung 51-inch plasma television mounted directly below a Sony EVI-D100 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 towards 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. This event type 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. 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.

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, even though the latter is highly salient perceptually, 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 we predict no asymmetry in dishabituation between the two conditions.

Results

Infants' average looking times by trial during the habituation and test phases of the experiment are plotted in Figure 1. To ensure that infants successfully habituated, we first compared the first three and last three trials of the habituation phase in a mixed-measures 2x2 ANOVA with condition (participant change vs. manner change) as a between-subject factor and habituation block (trial 1 through 3 vs. trial -3 through -1) as a within-subjects factor. We find a significant main effect of block ($F(1,30) = 155.89$, $p < .001$) but no main effect of condition ($F(1,30) = 0.03$, $p = 0.86$) and no interaction between the two ($F(1,30) = 0.03$, $p = 0.86$). That is, condition had no effect in the habituation phase, during which infants in both groups were exposed to the same stimulus (i.e., PICKING-UP videos).

To determine the extent to which infants treated the test videos as distinct from the habituation videos, we compared the final block of the habituation phase and the test phase in a

mixed measures 2x2 ANOVA with condition (participant change vs. manner change) as a between-subjects factor and phase (habituation vs. test) as a within-subjects factor. We find a significant main effect of phase ($F(1,30) = 76.76$, $p < .001$), indicating that infants dishabituated (i.e., recovered attention) upon reaching the test phase. Importantly, we find a significant interaction between condition and phase ($F(1,30) = 4.89$, $p < .05$), 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 test comparing looking times to the participant change and manner change videos during the test trials: infants in the participant change condition looked longer at test than infants in the manner change condition (Welch's $t(60.06) = 2.21$, $p < .05$).

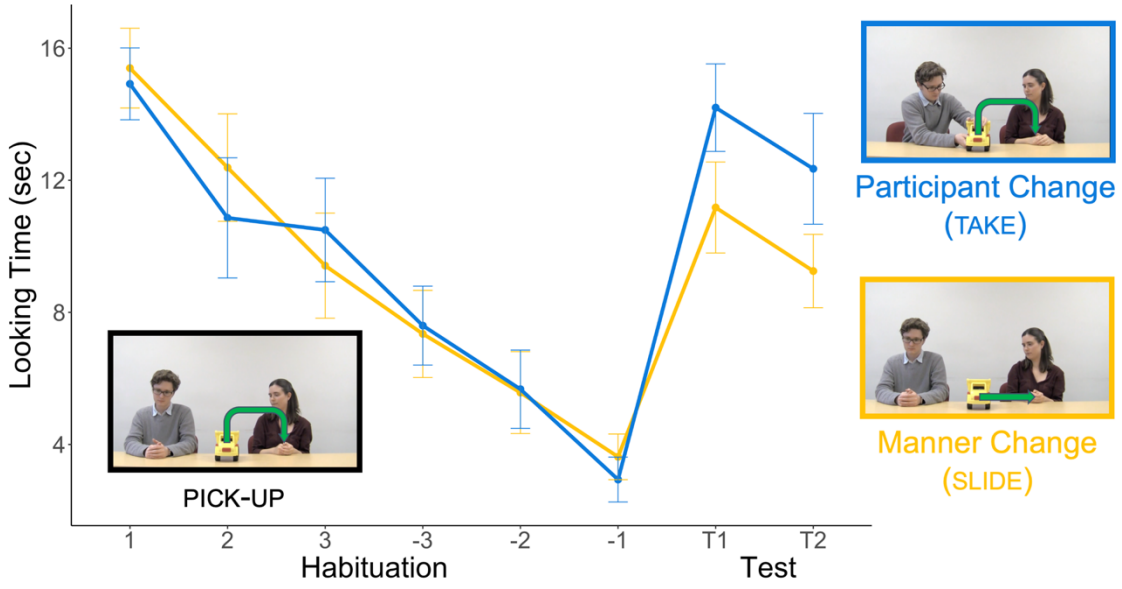

Figure 1. 10-month-olds' average looking time during the first block of the habituation phase (trials 1 through 3), the last block of the habituation phase (trials -3 through -1), and both trials of the test phase (trials T1 and T2). Both groups saw different tokens of the PICKING-UP video during the habituation phase. Infants in the "participant change" group saw the TAKING video during the test phase, whereas infants in the "manner change" group saw the SLIDING video. Error bars represent SEM.

Discussion

In sum, we find that infants registered a change from an event in which a girl moves a toy truck towards 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, indicating that they treated the boy's involvement as more noteworthy than a novel manner of motion (itself a large perceptual difference).

To be sure, infants also did distinguish the girl picking up the truck from the girl sliding the truck. But this should come as no surprise: they are different events, with different manners of motion. Not every change that infants can detect suggests a difference in the structure of the representations deployed. Infants can differentiate events that share conceptual structure but differ in content. To take another example, events of waving and events of jumping are represented under different concepts, with different entailments, but both concepts plausibly have only one privileged role, the agent.

The important point for our purposes is that infants reacted more strongly to the participant change (i.e., the difference between PICKING-UP and TAKING), despite the fact that the physical change here was smaller than that distinguishing the PICKING-UP from the SLIDING. It appears that infants viewed the 'taking' scene under a concept in which the boy filled a privileged role, one that was more psychologically potent than the manner of motion. To put the point another way, if the dishabituation observed between PICKING-UP and SLIDING reflects a difference in the content of the two event representations and the physical difference between two manners of motion, what explains the additional dishabituation observed between PICKING-UP and TAKING? On our view, this additional dishabituation reflects a difference in the structure

of the concepts. To return to an earlier analogy: we notice the difference between ‘37’ and ‘38’, which differ in content but share structure. But we likely experience an even greater difference between the base-10 ‘37’ and the base-2 ‘100110’ (=38), which differ in both content and structure.

All else being equal, then, the result reported above suggests that infants readily viewed our ‘taking’ scene under a conceptual representation in which the victim or source relation is explicit. Under the assumption that the taker (the agent) and thing taken (the patient) are also represented as participants, we might take this as evidence for a 3-participant conceptual structure, as in (4), where each participant role is again assigned its own label for clarity:

(4) TAKING-AGENT-PATIENT-SOURCE<e,x,y,z>

This is not to say that all events of taking are viewed under a 3-participant TAKING concept. When the source of a taking is particularly unresponsive, it may be ignored. For example, Tatone, Geraci & Csibra (2015) found that 12-month-old infants 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, Tatone & Csibra (2020), who argue for a similar GIVING/TAKING asymmetry in adults).

But as Tatone & 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 in terms of representing both under concepts that make three participants explicit. In our 'taking' event, the boy reacts to the truck being moved by retracting his hands and following the truck with his gaze. This seems to be enough of a social cue to encourage infants to represent the scene under a 3-participant TAKING concept that makes the source explicit.

Additional work might take further steps to explore these issues. But having some initial evidence for infants' representation of this particular 'taking' scene, we are now in a better position to determine which principles infants 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 deploy a more flexible strategy linking particular grammatical and participant relations? In Experiment 2, we pair this 'taking' stimulus scene with a novel verb in a sentence that has fewer than three arguments, and we ask what inferences 20-month-olds will draw about the meaning of that verb. This provides us with a test case for differentiating the One-to-One Matching strategy from more specific bootstrapping inferences based on thematic content.

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 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 towards 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 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 bootstrapping strategies 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 a 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) 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 not included in the final sample due to inattentiveness (6), equipment malfunction (1), or less than 80% English exposure (2). 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. 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.

		Sample Audio	**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?*		

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.

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 towards 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*, *shak*e, 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 a sub-event that has fewer numbers of

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 towards 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. 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. Thus, Thematic Linking predicts 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 towards either video. These average proportions of looks to TAKING are plotted by condition and window in Fig. 2. 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.

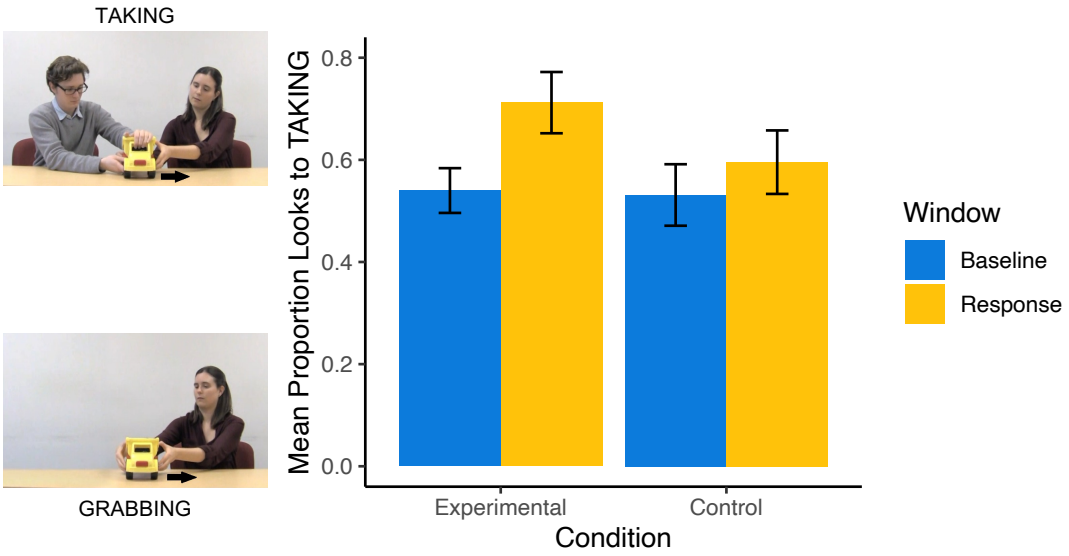

Figure 2. 20-month-olds' average proportion time spent looking to the TAKING video vs. the GRABBING video during the three seconds prior to novel verb offset (the baseline phase) and the three seconds following novel verb offset (the response phase). Infants in the experimental condition heard the novel verb in a transitive frame, and infants in the control condition heard the novel verb in an intransitive frame. Error bars represent SEM.

To assess the reliability of this pattern, we conducted a logistic mixed effects regression using the *lme4* package in R (Bates et al., 2015) with total frames spent looking to the TAKING video and to the MOVING video as the dependent variable. Note that this measure is no longer a proportion; instead, we use raw frame counts in order to retain information about how much total data an individual infant is contributing (Lidz et al., 2017). Fixed effects included window, condition, and their interaction. Factor contrasts were sum-coded. 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.

Significance testing was performed through log-likelihood ratio tests. Main effects were examined by comparing a model containing the fixed main effects against models that differed only in that they lacked the relevant fixed effect. These tests revealed a significant main effect of

window ($\chi^2(1) = 5.32$, $p < 0.05$), but no significant main effect of condition ($\chi^2(1) = 0.69$, $p = 0.40$). We then examined the window-by-condition interaction by comparing our full model against a model that differed only in that it lacked this interaction. Consistent with the visual pattern observed in Fig. 2, these tests revealed a significant 2-way interaction of window and condition ($\chi^2(1) = 3.98$, $p < 0.05$). Pairwise post-hoc comparisons using *Z*-tests, corrected with 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 ($Z = -3.07$, $p < 0.05$). In the control condition, there was no significant difference between the baseline and response windows ($Z = -0.34$, $p = 0.99$). 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.

Finally, to examine whether infants' verb vocabulary affected their behavior in this task, we fit another 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. Log-likelihood ratio tests again revealed a significant main effect of window ($\chi^2(1) = 96.51$, $p < 0.001$) and a significant interaction of window and condition ($\chi^2(1) = 45.24$, $p < 0.001$). The 2-way interaction of log verb vocabulary and window was also significant ($\chi^2(1) = 44.13$, $p < 0.001$), but no other significant interactions or main effects were found (all *p*s > 0.13). Importantly, there was no significant 3-way interaction of log verb vocabulary, window,

and condition ($\chi^2(1) = 1.64$, $p = 0.20$). 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, independent of condition, but it did not affect the different patterns of preference that we observe in the two conditions. These different preference patterns were exhibited to same extent in infants with low and high productive verb vocabulary.

Discussion

The results from Experiment 2 are inconsistent with the hypothesis that infants at 20 months expect the arguments in a clause to match one-to-one the participants that they perceive in an event. Infants 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), according to the results of Experiment 1. When asked to find *pimming* at test, infants in this condition showed above-baseline looking preferences for another token of the 3-participant TAKING event, preferring this scene over a 2-participant alternative in which the girl moved the truck towards herself in the same way, without the boy. This suggests that infants considered a 3-participant TAKING to be a better instance of *pimming* than a 2-participant GRABBING, even after familiarization with a 2-argument clause. This behavior is not predicted by One-to-One Matching: infants using this strategy should conclude that the transitive clause describes a sub-event, such as the girl's grabbing of the truck, in which perceived event participants match clause arguments in number. This prediction was not confirmed, suggesting that infants did not primarily rely on number-matching in this task.

Moreover, infants' behavior in the control condition rules out the possibility that infants' preferences in the experimental condition were driven by a familiarity bias, rather than by

sensitivity to clause syntax. Under this alternative account, infants may have preferred to map *pimming* to the 3-participant TAKING scene because this was the scene that was present during familiarization; the syntax of the familiarization sentence may have been ignored altogether. This would predict identical behavior for infants who heard transitive or intransitive syntax during familiarization. However, we found that infants who heard intransitive familiarization behaved differently from infants who heard transitive familiarization: they no longer showed an above-baseline preference for the TAKING video at test, but were instead indifferent about the two videos. This suggests that they viewed the GRABBING scene as an equally good instance of *pimming*, unlike infants in the experimental condition. Because the syntax of the familiarization clause led to different behavior in the two conditions, this indicates that infants used clause syntax 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.

Thus, these results show that the 20-month-olds in our task used syntax to draw inferences about verb meaning, but they did so in a way that was not predicted by One-to-One Matching. Their behavior is, however, predicted under the alternative Thematic Linking hypothesis. On this account, infants who heard *The girl pimmed the truck* during familiarization would link the subject to the agent of taking and the object to the patient. Because these relations align in the right way, this would enable them to infer that the transitive clause describes the entire 3-participant concept under which they readily viewed the TAKING scene. The sentence would not push them towards another 2-participant concept. Infants should thus conclude that *pimmings* can be 3-participant TAKINGS, and most likely are. However, they should not draw the same inference on the basis of hearing *The truck pimmed*. 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. Infants should thus 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 an account for infants' bootstrapping inferences at this age.

However, infants' indifference in the intransitive condition leaves open other ways of accounting for their behavior with this sentence type. On the account above, infants perceive a mismatch between their TAKING representation and an intransitive clause with *the truck* as subject, because they believe this sentence describes a change to the truck, independent of an agent. But it is also 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 types of events they might describe. If this is the case, the resulting confusion at test might manifest as apparent indifference. Prior experimental findings show 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), but further investigation is needed to assess whether 20-month-olds likewise have this ability, and can deploy it in our task.

Nonetheless, the current results demonstrate that infants' bootstrapping at 20 months is not primarily driven by the expectation of one-to-one matching between clause arguments and event participants. Instead, infants appear able to use a more flexible strategy to relate their sentence and scene representations, one that suggests sensitivity to differences in thematic content: a 3-participant TAKING can 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 when bootstrapping verb meanings.

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., 2016) 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 this representation fixed, 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 argument number, 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 the challenge. It is not unusual for a language to have morphologically unmarked intransitives whose predicate expresses a 2-participant event concept; and 3-participant event concepts find a happy home, we suggest, in ordinary transitive clauses. 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 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 (A. X. 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. 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: a learner who expects canonical clauses to have subjects may be able to use the distributions of noun phrases in intransitive clauses as evidence for the position of subjects vs. objects in the language (Perkins &

Hunter, in prep). 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. 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 NPs 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 "John loves Mary" as "L(j,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 be centered on a verb with two syntactic

"theta roles" that denotes a two-place function, this would have no logical 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.

(17999 words)

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.

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.

Arunachalam, S., & Waxman, S. R. (2010). Meaning from syntax: Evidence from 2-year-olds. *Cognition*, *114*(3), 442–446.

Baillargeon, R. (2004). Infants' physical world. *Current Directions in Psychological Science*, *13*(3), 89–94.

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, Martin, Bolker, Ben, & Walker, Steve. (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. (2015). Animacy and the acquisition of tough adjectives. *Language Acquisition*, *22*(1), 68–103.

Becker, M., Estigarribia, B., & Gylfadottir, D. (2012). Tough-adjectives are easy to learn. *Supplemental Proceedings of BUCLD*, *36*, 1–12.

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*, 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*, 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. (1976). Some linking regularities. In B. Levin & C. Tenny (Eds.), *On linking: Papers by Richard Carter*. Lexicon Project, MIT Center for Cognitive Science.

Christophe, A., Millotte, S., Bernal, S., & Lidz, J. (2008). Bootstrapping Lexical and Syntactic Acquisition. *Language and Speech*, *51*(1–2), 61–75.

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

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., 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.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 547–619.

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. *Developmental Psycholinguistics: On-Line Methods in Children's Language Processing*, *44*, 97.

Fillmore, C. J. (1968). The case for case. In E. Bach & R. Harms (Eds.), *Universals in Linguistic Theory*. 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.

Fisher, C., Gleitman, H., & Gleitman, L. R. (1991). On the semantic content of subcategorization frames. *Cognitive Psychology*, *23*(3), 331–392.

Fisher, C., Jin, K.-S., & Scott, R. M. (2019). The developmental origins of syntactic bootstrapping. *Topics in Cognitive Science*, *12*(1), 48–77.

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*. 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*(03), 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.

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature*, *415*(6873), 755–755.

Gertner, Y., & Fisher, C. (2012). Predicted errors in children's early sentence comprehension. *Cognition*, *124*(1), 85–94.

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science*, *17*(8), 684–691.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135–176.

Gleitman, L., & Papafragou, A. (2005). *Language and thought.* Cambridge University Press.

Gleitman, L. R. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55.

Gordon, P. (2003). The origin of argument structure in infant event representations. *Proceedings of the 26th Boston University Conference on Language Development*, 189–198.

Grimshaw, J. (1981). Form, function and the language acquisition device. In C. L. Baker & J. J. Mccarthy (Eds.), *The Logical Problem of Language Acquisition* (pp. 165–182). MIT Press.

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.

He, A. X. (2015). *Verb Learning Under Guidance*. http://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*, 1–22.

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. (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*.

Kratzer, A. (unpublished m.s.). The event argument and the semantics of verbs. In *The Event Argument*.

Kuhlmeier, V., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological Science*, *14*(5), 402–408.

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.

Levin, B., & Hovav, M. R. (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.

Lidz, J., & Gleitman, L. R. (2004). Argument structure and the child's contribution to language learning. *Trends in Cognitive Sciences*, *8*(4), 157–161.

Lidz, J., White, A. S., & Baier, R. (2017). The role of incremental parsing in syntactically conditioned word learning. *Cognitive Psychology*, *97*, 62–78.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman and Company.

Meltzoff, A. N. (1988). Infant imitation after a 1-week delay: Long-term memory for novel acts and multiple stimuli. *Developmental Psychology*, *24*(4), 470.

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.

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. https://books.google.com/books?hl=en&lr=&id=RWeYAgAAQBAJ&oi=fnd&pg=PA1&dq=Morgan+%26+Demuth,+1996&ots=or8wK2wcFZ&sig=BcWTt6QncO73md9g35I1FjUb8CE

Muentener, P., & Carey, S. (2010). Infants' causal representations of state change events. *Cognitive Psychology*, *61*(2), 63–86.

Naigles, L. R. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, *17*(2), 357–374.

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.

Pearl, L., & Sprouse, J. (2019). Comparing solutions to the linking problem using an integrated quantitative framework of language acquisition. *Language*, *95*(4).

Perkins, L., & Hunter, T. (in prep). *Word order acquisition from immature representations: A model for noise-tolerant learning of deterministic grammars*.

Perkins, L., & Lidz, J. (2020). Filler-gap dependency comprehension at 15 months: The role of vocabulary. *Language Acquisition*, *27*(1), 98–115.

Perkins, L., & Lidz, J. (2021). 18-month-old infants represent non-local syntactic dependencies. *Proceedings of the National Academy of Sciences*, *118*(41), e2026469118.

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.

Pinker, S. (1984). *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.

Pinker, S. (1989). *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge, MA: MIT Press.

Pinker, S. (1994). How could a child use verb syntax to learn verb semantics? *Lingua*, *92*, 377–410.

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*.

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.

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.

Snedeker, J., & Gleitman, L. (2004). Why it is hard to label our concepts. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a Lexicon*. MIT Press.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96.

Syrett, K., Musolino, J., & Gelman, R. (2012). How can syntax support number word acquisition? *Language Learning and Development*, *8*(2), 146–176.

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*.

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.

Tatone, D., Hernik, M., & Csibra, G. (2021). Facilitation of object encoding in infants by the observation of giving. *Scientific Reports*, *11*(1), 1–9.

Wagner, L., & Carey, S. (2005). 12-month-old infants represent probable endings of motion events. *Infancy*, *7*(1), 73–83.

Waxman, S. R. (1999). Specifying the scope of 13-month-olds' expectations for novel words. *Cognition*, *70*(3), B35–B50.

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.

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.

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.

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.

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.

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.

White, A. S., Hacquard, V., & Lidz, J. (2016). Main clause syntax and the labeling problem in syntactic bootstrapping. *Semantics in Acquisition. Trends in Language Acquisition Research (TiLAR). Amsterdam: John Benjamins Publishing Company*.

Williams, A. (2015). *Arguments in syntax and semantics*. Cambridge University Press.

Williams, A. (2021). Events in semantics. In *Cambridge Handbook of the Philosophy of Language*. 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*, *33*(33).

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1–34.

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.

Younger, B. A., & Cohen, L. B. (1985). How infants form categories. In *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.

Yuan, S., Fisher, C., & Snedeker, J. (2012). Counting the nouns: Simple structural cues to verb meaning. *Child Development*, *83*(4), 1382–1399.

Appendix

Character Familiarization	**Training: *Shake***
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: **Control Condition (Intransitive)**	**Experimental Trial:** **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 A1. Audio stimuli, Experiment 2.

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

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