The Power of Ignoring: Filtering Input for Argument Structure Acquisition

Laurel Perkins[1], Naomi H. Feldman[2,3], Jeffrey Lidz[2]

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

[2]Department of Linguistics, University of Maryland

[3]Institute for Advanced Computer Studies, University of Maryland

Author Note

Address for correspondence:

Laurel Perkins

3125 Campbell Hall

Los Angeles, CA 90025

perkinsl@ucla.edu

Abstract

Learning in any domain depends on how the data for learning are represented. In the domain of language acquisition, children's representations of the speech they hear determine what generalizations they can draw about their target grammar. But these input representations change over development as a function of children's developing linguistic knowledge, and may be incomplete or inaccurate when children lack the knowledge to parse their input veridically. How does learning succeed in the face of potentially misleading data?

We address this issue using the case study of "non-basic" clauses in verb learning. A young infant hearing *What did Amy fix?* might not recognize that *what* stands in for the direct object of *fix*, and might think that *fix* is occurring without a direct object. We follow a previous proposal that children might filter non-basic clauses out of the data for learning verb argument structure, but offer a new approach. Instead of assuming that children identify the data to filter in advance, we demonstrate computationally that it is possible for learners to infer a filter on their input without knowing which clauses are non-basic. We instantiate a learner that considers the possibility that it mis-parses some of the sentences it hears, and learns to filter out those parsing errors in order to correctly infer transitivity for the majority of 50 frequent verbs in child-directed speech. Our learner offers a novel solution to the problem of learning from immature input representations: learners may be able to avoid drawing faulty inferences from misleading data by identifying a filter on their input, without knowing in advance what needs to be filtered.

Keywords: Language acquisition, verb learning, argument structure, bootstrapping, computational modeling, Bayesian inference

1 Introduction

Learning involves incrementally building on prior knowledge. This is true in language acquisition just as in other forms of learning: a child who can't count cannot learn arithmetic, and a child who can't identify the category 'verb' cannot learn whether her language has verb raising. Linguistic theories since Chomsky (1965) have typically abstracted away from time and resource constraints, idealizing language acquisition as an instantaneous process that maps the input onto a grammar. Some modern approaches make the same idealization in asking whether the data as a whole support grammar selection (e.g., Perfors, Tenenbaum, & Regier, 2006; Yang, 2002). But while enabling insights into language learnability at a global level, these approaches abstract away from another important dimension of the learning problem: how learners perceive and use their input, and how this changes as they learn their language.

This paper investigates a puzzle that arises from incorporating development into a model of grammar acquisition. At any given point in development, the way that children perceive their input depends on their current knowledge of their language, which they use to assign structure and meaning to the speech that they hear. These input representations change as children's linguistic knowledge develops, and determine what further inferences children can draw about their target grammar. Learning cannot wait until children can veridically parse all of their input, or there would be nothing further to learn; instead, children must learn from the immature parses that they can assign to their input at each stage of development (Fodor, 1998; Valian, 1990). How do learners avoid being misled if their input representations are incomplete or inaccurate?

Our case study is the role of transitivity in verb learning. At very early stages in grammatical development, learners use verbs' distributions in transitive and intransitive clauses to draw inferences about verb meanings and argument structure (Fisher, Gertner, Scott, & Yuan, 2010; Lidz, White, & Baier, 2017). But accurately perceiving those distributions is not trivial, as transitive and intransitive clauses can be realized in variable

ways within a language and cross-linguistically. The arguments in "basic" English clause types like (1) and (2) might be easier to recognize than those in "non-basic" clause types that do not follow the language's canonical word order, like (3):

(1) John ate a sandwich. Amy fixed her bicycle.

(2) John ate. (*Amy fixed.)

(3) What did John eat? What did Amy fix?

If a child knows that English has canonical subject-verb-object word order, she could recognize that the sentences in (1) contain both subjects and objects, and the sentences in (2) contain only subjects. These data could lead her to conclude that *fix* is obligatorily transitive whereas *eat* can alternate between transitive and intransitive uses, which has implications for what those verbs might mean. But transitivity might be harder to recognize in the wh-object questions in (3), in which a fronted argument (*what*) stands in a non-local dependency with the verb, and acts as the verb's object even though it does not surface in canonical direct object position. These data may be misleading for a child who has not yet learned how to identify wh-dependencies in her language, and does not know that *what* is a wh-word. She might note the absence of a direct object after the verb and perceive the sentences in (3) as intransitive, mistakenly concluding that *fix* can alternate just like *eat*.

One solution to this problem proposes that learners somehow "filter out" non-basic clauses like wh-object questions early in language acquisition. Under this approach, young children avoid learning about basic argument structure, clause structure, and verb meanings from sentences that do not follow the canonical word order of their language, because these sentences obscure the systematic relations between syntax and semantics that are useful for learning (Gleitman, 1990; Lidz & Gleitman, 2004a, 2004b; Pinker, 1984, 1989). This approach has implicitly assumed that learners know which sentences to filter out, but the mechanism by which they identify these sentences has not yet been established. Furthermore, learning to identify argument displacement in non-basic clauses would seem to depend on

knowing some core argument structure properties of the language: learning that *what* is the object of *fix* in a wh-question like (3) arguably depends on knowing that *fix* takes a direct object (Gagliardi, Mease, & Lidz, 2016; Perkins & Lidz, 2020). Thus, an apparent paradox arises. Learning basic verb transitivity, a first step in the acquisition of argument structure, may require filtering out data from non-basic clauses. But identifying which clauses are non-basic may require already knowing which verbs are transitive. Empirical evidence suggests that learners face this paradox in their second year of life. Findings from behavioral studies show that verb transitivity knowledge develops in tandem with infants' ability to identify common non-basic clause types like wh-dependencies, before they turn two years old (Gagliardi et al., 2016; Lidz et al., 2017; Perkins & Lidz, 2020; Perkins, 2019; Seidl, Hollich, & Jusczyk, 2003).

Here, we resolve this apparent paradox computationally. We present a Bayesian model that learns to filter its input to infer verb transitivity, without knowing what types of sentences it should filter out. Our model does so under the assumption that it occasionally parses sentences erroneously, and it learns how much of its parses to trust and how much it should treat as noise for the purposes of verb learning. This allows the learner to avoid drawing faulty inferences from non-basic clauses, without having to know which clauses are non-basic. In simulations on child-directed speech, we show that our model learns appropriate parameters for filtering its input in order to accurately categorize the majority of frequent transitive, intransitive, and alternating action verbs. We thus provide a model for the first steps of argument structure acquisition that have been attested in infancy, demonstrating how those steps of learning could take place before non-basic clause acquisition is complete. In doing so, we propose a new solution to the problem of learning from input that a learner cannot parse veridically. It may be possible for learners to avoid drawing faulty inferences from misleading data by identifying a filter on their input, without knowing in advance what needs to be filtered.

2 Non-Basic Clauses in Verb Learning

Non-basic clauses are problematic for theories of learning that rely on systematic relations between verbs' syntactic properties and their meanings, e.g. semantic and syntactic bootstrapping (Fisher et al., 2010; Gleitman, 1990; Grimshaw, 1981; Landau & Gleitman, 1985; Lasnik, 1989; Pinker, 1984, 1989). It is not trivial to identify which sentences have undergone particular transformations, and learners who do not yet know the surface signals for these transformations might mis-perceive the structure of non-basic clause types in their input[1]. If so, this will disrupt learners' attempts to put the syntactic environments in which verbs occur into correspondence with conceptual categories of events they perceive in the world. Therefore, syntactic and semantic bootstrapping theories have traditionally assumed that learners at early stages of grammatical development have some way to avoid learning from clauses that are non-basic.

2.1 Semantic and Syntactic Bootstrapping

Semantic and syntactic bootstrapping rely on correspondence relations between linguistic and conceptual structure. If the syntactic environments in which verbs distribute are related in a systematic way to conceptual categories of events they describe, then learners can use evidence about one of these properties (syntactic or conceptual) to draw inferences about the other. In semantic bootstrapping, a child who represents an event under a particular conceptual structure might be able to use these correspondence relations to draw inferences about the syntactic structure of the clause describing that event (Grimshaw, 1981; Pinker, 1989, 1984). For example, a child who perceives an event as involving an agent and a patient, and furthermore knows that subjects of active transitive clauses tend to name agents and objects tend to name patients, could then infer which argument is the subject and which

[1]Note that this problem is not unique to transformation-based grammatical theories. Under theories in which transitive clauses, wh-object questions, and passives are separate "constructions" (Fillmore, Kay, & O'connor, 1988; Goldberg, 1995; Langacker, 1999), the learner must still ultimately recognize that only verbs that occur in transitives can also occur in wh-object questions and passives. Whether this is encoded transformationally or via a construction hierarchy, the same logical problem holds.

is the object in a clause describing that event. Conversely, in syntactic bootstrapping, a child who represents a clause under a particular linguistic structure might be able to use these correspondence relations in the opposite direction to draw inferences about which event the clause describes (Fisher et al., 2010; Gleitman, 1990; Landau & Gleitman, 1985; Lasnik, 1989). For example, a child who hears an unknown verb in a clause that she represents as transitive could then infer that this clause describes an event she perceives as having an agent and a patient, allowing her to narrow down the range of events that the new verb might describe.

Experimental tests of verb learning find evidence that learners in their second year of life are beginning to use these meaning-distribution correspondence relations, particularly those that pertain to transitivity. In preferential looking tasks, English-learning infants as young as 17 months can use the canonical subject-verb-object word order of English to identify that the individual named by the subject of a transitive clause is the agent of an event, and the individual named by the object is the patient[2] (Hirsh-Pasek & Golinkoff, 1996; Gertner, Fisher, & Eisengart, 2006). By the age of 19 months, infants reliably infer a causative meaning for a novel verb in a transitive vs. an intransitive clause, and do so under the right circumstances at 15 months as well (Arunachalam & Waxman, 2010; Arunachalam, Escovar, Hansen, & Waxman, 2013; Jin & Fisher, 2014; Messenger, Yuan, & Fisher, 2015; Naigles, 1990; Yuan & Fisher, 2009; Yuan, Fisher, & Snedeker, 2012). Children draw even finer-grained inferences on the basis of hearing novel verbs participate in particular transitive-intransitive alternations. The subject of an intransitive clause can name either an agent (e.g. *John baked*) or a patient (e.g. *The bread rose*). How a verb distributes in intransitive clauses is related to its meaning: intransitives whose subjects are agents tend to describe activities of those agents, whereas intransitives whose subjects are patients tend to describe changes undergone by those patients (Fillmore, 1968, 1970; Levin & Hovav, 2005;

[2]We note that these data do not tell us how the categories 'subject' and 'object' are represented by infants at this age— that is, whether infants represent arguments within a hierarchical clause structure, or merely encode their linear order (Fisher, 1996).

Williams, 2015). Another line of experimental work has found that children by the age of 2 are sensitive to these distinctions (Bunger & Lidz, 2004, 2008; Naigles, 1996; Scott & Fisher, 2009).

In summary, bootstrapping allows young learners to draw inferences about grammar and meaning by relating syntactic representations of subjects and objects in sentences to conceptual representations of events. But these inferences, whether they are syntactic bootstrapping or semantic bootstrapping inferences, depend on learners recognizing subjects and objects when they are present, and may fail if other linguistic properties interfere with learners' abilities to recognize those core clause arguments.

2.2 The Problem of Non-Basic Clauses

Both bootstrapping theories acknowledge that the correspondence relations between syntax and meaning only hold probabilistically, and may be obscured when they interact with other grammatical properties of the language (Gleitman, 1990; Lidz & Gleitman, 2004a, 2004b; Pinker, 1984, 1989). This problem was first noted by Pinker in his earliest work on semantic bootstrapping, following Keenan (1976):

> One must place an important proviso, however, on the use of semantic information to infer the presence of syntactic symbols, especially grammatical relations. Keenan argues that the semantic properties of subjecthood hold only in what he calls "basic sentences": roughly, those that are simple, active, affirmative, declarative, pragmatically neutral, and minimally presuppositional. In nonbasic sentences, these properties may not hold. In English passives, for example, agents can be oblique objects and patients subjects, and in stylistically varied or contextually dependent sentences the agent can be found in nonsubject positions (e.g., *eats a lot of pizza, that guy*). Thus one must have the child not draw conclusions about grammatical relations from nonbasic sentences (Pinker, 1984).

To appreciate the full extent of Pinker's problem, let us consider the case of the

wh-object question in (3), repeated here as (4), as well as other non-basic clause types such as relative clauses (5) and passives (6).

(4) What did Amy fix?

(5) I like the bicycle that Amy fixed.

(6) The bicycle was fixed (by Amy).

In each of these examples, a syntactic transformation has applied such that the argument acting as the object of the verb no longer surfaces in canonical object position. If a child is not aware of these transformations, she may be misled when she relates the linguistic structure she (mis-)perceives in these clauses with her conceptual representations of events. For example, a semantic bootstrapper who takes (6) to be a description of an event in which she perceives Amy to be the agent and the bicycle to be the patient might construe "Amy" as the subject and "the bicycle" as the object, resulting in a parse that is not only erroneous but also implies that English has object-verb-subject (OVS) word order. Likewise, the fronted arguments in (4) and (5), if recognized as arguments, might be taken as evidence for optional OSV word order in English rather than as evidence for the wh-movement that actually produced this non-canonical word order. And if these phrases are not recognized as arguments of *fix*, a variety of other inaccurate parses would be available for these sentences: perhaps English allows syntactic null objects, or perhaps *fix* can take an implicit object like *eat*. This could lead to faulty inferences about the syntactic properties of particular verbs and of the grammatical properties of the target language.

Conversely, a syntactic bootstrapper who is not aware of the transformations in these sentences may draw faulty inferences about which events in the world they describe. Because direct objects are not realized in their canonical post-verbal position, a child may not recognize that these clauses are underlyingly transitive, and thus may not infer that they describe causative events. In this case, an event in which she perceives Amy to be the agent and the bicycle to be the patient may no longer count for her as a possible "fixing." The

problem is not necessarily solved as soon as she observes *fix* in a basic clause that she can recognize as transitive. In that case, she may infer that *fix* belongs to some class of verbs that can alternate between transitive and intransitive uses, like *eat* or *rise*, leading to inaccurate inferences about both its syntactic and semantic properties.[3]

2.3 Empirical Evidence

Empirical evidence shows that learners encounter this problem very early in grammatical development. Non-basic clauses are prevalent in the input to infants. Although Pinker's initial survey of infant-directed speech found very few instances of passives (Pinker, 1984), other studies find relatively high rates of other non-basic clause types. In particular, English-learning children hear a large number of wh-questions before their second birthday (around 15% of their total input), the majority of which contain non-canonical word orders (Cameron-Faulkner, Lieven, & Tomasello, 2003; Newport, Gleitman, & Gleitman, 1977; Stromswold, 1995). Identifying the structure of these clause types requires knowing how particular transformations are realized in the target language. For example, in order to identify the structure of wh-object questions like (4), a child must detect that a fronted argument (*what*) stands in relation to a verb (*fix*) that needs an object and is locally missing one. But this requires the child to know that *what* is an argument, even though it is a functional element that does not distribute like other arguments in the language. The child would also need to know that *fix* needs an object, and is not merely being used intransitively.

Furthermore, experimental findings suggest that infants' abilities to identify the structure of these common non-basic clause types develops in tandem with basic argument structure knowledge. Infants as young as 15 and 16 months old show sensitivity to verb

[3]Note that this problem is not solved under the hypothesis that verb meanings play a role in acquiring verb alternation properties (Pinker, 1989). A learner who believes that a verb describes an event with an agent and a patient still cannot be certain whether the verb will syntactically alternate (although subtler conceptual correlates may be informative; see Resnik (1996)). Eatings always involve an eater and a thing eaten, and fixings always involve a fixer and a thing fixed, but *eat* can freely drop its object and *fix* cannot. This means that a learner who fails to recognize the displaced objects in (4-6) now has a choice: if she knows that fixings always involve something fixed, she might be suspicious that these sentences have objects after all, or she might conclude that *fix* allows object-drop just like *eat*.

transitivity. Jin and Fisher (2014) found that 15-month-olds are able to draw inferences about the meaning of a novel verb on the basis of hearing it in a transitive frame, and Lidz et al. (2017) found that high-vocabulary 16-month-olds predicted an upcoming direct object for a known transitive verb during online sentence processing. In a listening-time study, Perkins (2019) found that 15-month-olds detected when common transitive verbs were missing a direct object, differentiating between these ungrammatical uses and grammatical uses in transitive frames.

However, infants' wh-dependency knowledge at 15 months appears fragile. One early preferential looking study found that 15-month-olds were able to comprehend subject but not object wh-questions (Seidl et al., 2003). In two additional studies that found apparent success with object questions at this age (Gagliardi et al., 2016; Perkins & Lidz, 2020), the authors argued that 15-month-olds' performance was not due to an adult-like representation of the wh-dependencies in these sentences, but rather to an interpretive heuristic based on verb knowledge. If infants know that a verb like *bump* requires a direct object, then a question like *Which dog did the cat bump?* might lead them to look towards the patient of bumping by the cat, even if they don't syntactically represent the fronted wh-phrase as that object.

In support of this account, the listening-time study in Perkins (2019) found that 15-month-olds did not differentiate between a locally missing object in a basic declarative clause vs. an object wh-question. Infants at this age responded in the same way to all sentences with locally missing objects (e.g. **A dog! The cat should bump / Which dog should the cat bump?*), compared to sentences with objects after the verb (e.g. *A dog! The cat should bump him / *Which dog should the cat bump him?*). This suggests that 15-month-olds were aware that these transitive verbs needed objects, but were unaware that the wh-phrase satisfies that requirement non-locally in a wh-question. By contrast, 18-month-olds did show this awareness: they showed opposite preferences for local objects in declaratives and wh-questions, listening longer to grammatical sentences of each type. These results suggest

that infants represent the wh-phrase as a non-local object of the verb at 18 months, but not earlier. Additional results show that infants begin to produce wh-questions in their own speech by 20 months (Rowland, Pine, Lieven, & Theakston, 2003; Stromswold, 1995) and reliably comprehend them in preferential looking tasks at this age (Gagliardi et al., 2016; Seidl et al., 2003).

In summary, the current experimental evidence points towards the following developmental trajectory. Basic verb transitivity knowledge appears to develop early, at 15-16 months for English learners, and emerges before infants represent a fronted wh-phrase as an argument in a wh-question, at 18-20 months. This implies that infants must have a way to begin learning argument structure and basic clause structure even before they can parse some of the most common non-basic clause types in their input— and even though those clause types provide very misleading data for bootstrapping.

2.4 Filtering

The solution proposed in the bootstrapping literature is that learners' input must be filtered in such a way as to boost the signal from basic clauses, in which core arguments will be easier to identify and correspondence relations between syntax and meaning will hold more reliably (Gleitman, 1990; Lidz & Gleitman, 2004a, 2004b; Pinker, 1984, 1989). In other words, non-basic clauses are somehow filtered out of the data that young children use to bootstrap basic argument structure and clause structure.

Pinker (1984) proposes two ways that this filtering might happen: either parents might do the filtering and avoid producing these sentences in their children's presence, or children might internally filter these sentences themselves. Parental filtering does not seem to occur, as evidenced by the high rate of wh-questions in speech to young infants (Cameron-Faulkner et al., 2003; Newport et al., 1977; Stromswold, 1995). The second logical solution is for children to filter out non-basic clauses themselves. This approach implicitly assumes that children know which sentences to filter out. But this solution risks being circular. Learners

need to filter non-basic clauses in order to learn argument structure, but identifying the structure of non-basic clauses would seem to depend on knowing some core argument structure properties of the language. And the empirical evidence suggests that these two phenomena are acquired in tandem, with verb transitivity knowledge emerging a few months before learners recognize displaced arguments in wh-questions. How can learners identify non-basic clauses in order to filter them, if they do not know yet what argument displacement looks like in their language, and are still in the process of bootstrapping such basic grammatical properties as verb transitivity?

Pinker (1984, 1989) argues that this circularity can be avoided if children can use non-syntactic cues to flag certain utterances as likely to contain non-basic clauses, without recognizing the structure of those clauses. These cues might include "special intonation, extra marking of the verb, presuppositions set up by the preceding discourse or the context, nonlinguistic signals of the interrogative or negative illocutionary force of an utterance" (Pinker, 1984). The challenge with this solution is identifying how learners know which cues to use. Attempting to define the criteria by which children should filter their input creates its own learning problem: this introduces a new set of categories which the learner must know to track, and which in many cases may be far from transparent (Gleitman, 1990).

One might imagine the following solution to the circularity problem: perhaps learners acquire non-basic clause syntax and verb argument structure by attempting to learn both of these phenomena at the same time. This simultaneous learning hypothesis may be in principle possible. But we choose to model a different hypothesis instead, one that allows learning to take place in incremental steps over development, and thus provides an account for the empirically observed developmental trajectory for these phenomena. Our solution does not require learners to know the criteria for identifying non-basic clauses in order to learn verbs. One way of thinking about this is that it provides learners with a way of arriving at a fruitful starting point from which a subsequent joint learning process might proceed—an initial bootstrap into the system.

We focus on the most basic and earliest-acquired verb argument-taking properties, transitivity, and ask how this property is acquired by infants who do not yet know how to parse non-basic clauses in their input. We propose that young learners implicitly assume that they will not accurately parse everything they hear, and expect that their data will contain a certain amount of noise: erroneous parses that they shouldn't trust for the purposes of verb learning. Children might be able to learn the right way to filter erroneous parses out of their input in order to solve a particular learning problem– in this case, jointly inferring verb transitivity along with how much of their data to trust in making that inference. Crucially, this solution doesn't require learners to know where those errors came from, thereby sidestepping the problem of which cues learners should track for identifying non-basic clauses. Under our approach, children might filter non-basic clauses from the data they use for verb learning without knowing that they are non-basic clauses. This will allow them to use relations between syntax and meaning to bootstrap into the target grammatical system, even though they do not yet know when those relations are masked by other grammatical properties of the language.

2.5 Computational Models of Verb Learning

We adopt a Bayesian framework, in which a learner observes a data pattern and infers the probability of some properties of the system that may have generated that data. This framework conveniently allows us to specify the alternative systems (verb transitivity properties vs. erroneous parses) that our learner considers for the verb distributions it observes.

Our model follows previous Bayesian approaches to argument structure acquisition (Alishahi & Stevenson, 2008; Barak, Fazly, & Stevenson, 2014; Parisien & Stevenson, 2010; Perfors, Tenenbaum, & Wonnacott, 2010), but considers a different problem than the one explored in that literature. The goal of the learners in Alishahi and Stevenson (2008) and Perfors et al. (2010) is to identify the full set of verb classes that exist in the language, and

how verbs in those classes generalize across syntactic frames. These papers aim to provide an account for an acquisition phenomenon that arises in preschool-aged children, who sometimes over-generalize verbs across argument structures that they do not actually participate in. This behavior occurs in children at a later stage of development than the infant bootstrappers we are modelling in this paper. Verb over-generalization is the output of several logically independent steps of learning: (1) perceiving how verbs distribute in particular syntactic frames; (2) performing an initial classification of verbs according to their argument-taking properties, e.g., as one-, two-, or three-place predicates; and finally (3) identifying how productively verbs in a class can generalize across different types of argument structures, e.g., from the prepositional dative to the double-object dative. The primary focus of prior models is the third step of learning, but we are concerned with the earlier processes involved in the first two steps. In particular, we ask how learners are able to establish a veridical percept of verbs' distributions with subjects and objects, when they may not have the linguistic knowledge to reliably identify these core syntactic arguments in non-basic clauses.

This question has not yet been answered in previous models of argument structure acquisition, in which a learner's ability to veridically represent the input has been largely assumed. Alishahi and Stevenson (2008) acknowledge that this assumption is most likely unrealistic, and simulate noise in their learner's syntactic representations by randomly removing some of the distributional features that it learns from. Yet the "noise" faced by a learner in real life is not random. As the authors note, "A more accurate approach must be based on careful study of the types of noise that can be observed in child-directed data, and their relative frequency" (Alishahi & Stevenson, 2008). This invites us to consider the ways in which learners might mis-perceive the data in their input, and how a learner can avoid being misled by that data when identifying a verb's syntactic distribution in the language.

Our approach differs from these prior verb learners in another important way. Rather than modelling the simultaneous acquisition of all verb classes, we focus on only verb

transitivity as the earliest-attested form of argument structure knowledge in infancy, and arguably the most basic. This allows us to explicitly model a particular developmental stage suggested by the empirical literature, which learners transit on their way to acquiring the full argument structure system of their language. That is, we model development by breaking a large acquisition problem into smaller steps. We ask how learners first identify the core argument-taking properties of verbs— their distributions with subjects and objects— in order to provide a scaffold for further inferences about the target grammar, including the finer-grained distributional classes and alternations that verbs participate in.

Other previous computational models have investigated how learners might benefit from simultaneously making use of semantic information during this process. These models ask how learners might use conceptual structure to identify the core grammatical rules and word order properties of the language, and then use syntactic representations to infer the meaning of words and utterances (Abend, Kwiatkowski, Smith, Goldwater, & Steedman, 2017; Kwiatkowski, Goldwater, Zettlemoyer, & Steedman, 2012; Maurits, Perfors, & Navarro, 2009). While shedding light on how semantic and syntactic bootstrapping might proceed in tandem, these models still presuppose the step of learning that we are concerned with in this paper: how a learner gains access to accurate representations to form the basis of these bootstrapping inferences. As prior work as noted, the noise introduced by non-basic clauses types is equally disruptive for both types of bootstrapping, semantic or syntactic (Pinker, 1984, 1989; Gleitman, 1990). We focus here on the learner's syntactic percept, but in doing so, we do not deny that it might be helpful to make use of conceptual information as well. Our goal is simply to ask how far a learner could get in identifying verb transitivity on the basis of distributional information, when those distributions may not be accurately perceived. In order to isolate this distributional signal for bootstrapping, we therefore set aside the question of how conceptual information could be accessed or used in this process, a question we will return to later in the discussion.

In the experiments below, we test the computational feasibility of our proposed

solution: whether a learner could, in principle, jointly infer verb transitivity along with the parameters for filtering errorful sentence representations from the data it uses for learning. In Simulation 1, we demonstrate that a learner can accomplish this joint inference on the basis of the syntactic distributions of frequent English action verbs in child-directed speech. Our learner performs this inference using only rates of overt direct objects after verbs, and does not condition on any other utterance features, such as wh-words, prosody, or extra-linguistic discourse context; it succeeds even though it cannot distinguish object wh-questions from basic intransitive clauses. In Simulation 2, we ask how much the learner's performance in Simulation 1 depended on its *a priori* assumption that transitive, intransitive, and alternating verbs are equally likely. We show that our learner performs no better when it assumes these categories will occur in the proportions in which they actually do occur in child-directed English. However, it does not differentiate transitivity categories as well when it is extremely biased towards the alternating class, showing that the deterministic categories must be weighted sufficiently in the model's hypothesis space in order to be identified in its input. Thus, we provide a proof of concept that a child may be able to filter non-basic clauses from her input in order to correctly identify verb transitivity, without knowing in advance which clauses are non-basic. This inference requires prior knowledge that verbs might be transitive or intransitive, but does not require specific knowledge about the frequency of those transitivity categories in the learner's target language.

3 Model

We present a Bayesian model that learns how to filter its input in order to infer verb transitivity. The learner performs this inference only on the basis of observing how verbs distribute with and without direct objects, and does not use any other syntactic or non-syntactic cues to identify its filter. Instead, the learner assumes that some of its parses are not trustworthy sources of information for learning its language, because it does not have enough linguistic knowledge to accurately parse every sentence in its input. The learner

infers the right way to filter erroneous parses out of the data it uses for verb learning, without knowing why those parses were erroneous.

In this section, we first specify the generative model, which encodes the learner's assumptions about how its direct object observations are generated. Then, we specify how the learner jointly infers verb transitivity along with the parameters for filtering its input, given its data. In the following sections, we present simulations demonstrating that this joint inference is successful when tested on child-directed speech. We do not claim that the Bayesian inference performed by our model represents the exact algorithms performed by child learners. Although there is substantial literature on young children's statistical inference capabilities (Gomez & Gerken, 2000), this model is intended only as a proof of concept that such joint inference is possible. However, although this model may not provide a realistic implementation of the inference process that children use, it provides a more realistic account than previous models of the steps of learning involved in bootstrapping: specifically, how learners establish a veridical percept of verbs' syntactic distributions, in order to enable further bootstrapping inferences.

3.1 Generative Model

A generative model represents a learner's assumptions about the processes that generated its observed data. In our case, the observed data are counts of direct objects with particular verbs, as the learner represents them; specifically, the learner tracks how frequently it represents an overt direct object or no overt direct object following the verb. It assumes that there are two reasons why it might observe direct objects or no direct objects. On one hand, the transitivity of the verb determines whether it always, never, or sometimes takes a direct object. This means that the rate of direct objects following the verb gives the learner evidence for inferring whether the verb is transitive, intransitive, or alternating. But on the other hand, the learner might also mis-perceive whether a direct object is present, because it lacks the grammatical knowledge to identify the full structure of some sentences in

its input. If this is the case, some of the observed data points might not reflect the true transitivity of the verb and should be filtered from the data that the learner uses to infer transitivity. Thus, there is some probability of error in the learner's direct object observations, and our learner infers two parameters for filtering this error: how frequently mis-parses of sentences occur, and whether the learner is more likely to miss a direct object that is underlyingly present or mistake another constituent for a direct object.

Figure 1 provides the graphical model for our learner. The model's observations of direct objects or no direct objects are formalized as the Bernoulli random variable X. Each $X^{(v)}$ represents an observation from a sentence containing verb v in the model's input, with a value of 1 if the sentence contains a direct object and 0 if it does not. These observations of direct objects can be generated by two processes: the transitivity of verb v, represented by the variables T and θ in the upper half of the model, or an erroneous parse of the sentence, represented by the variables e, ϵ, and δ in the lower half of the model. We will describe each of these processes in turn.

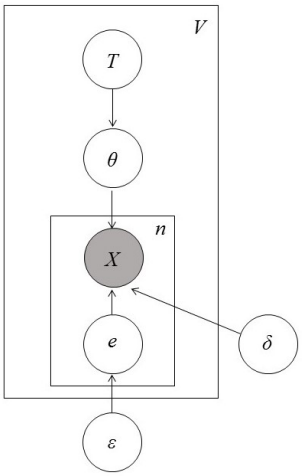

Figure 1. Graphical Model

In the upper half of the model, each $X^{(v)}$ is conditioned on the parameter $\theta^{(v)}$, a continuous random variable defined for values from 0 to 1 inclusive. This parameter controls how frequently a verb v will be used with a direct object: the learner assumes that for every observation $X^{(v)}$, a biased coin is flipped to determine whether the sentence contains a direct

object, with probability $\theta^{(v)}$, or does not, with probability $1 - \theta^{(v)}$. The parameter $\theta^{(v)}$ is conditioned on the variable $T^{(v)}$, which represents the transitivity of verb v. T is a discrete random variable that can take on three values, corresponding to transitive, intransitive, and alternating verbs. Each of these values determines a different distribution over θ. For the transitive category of T, θ always equals 1: the verb should always occur with a direct object. For the intransitive category, θ always equals 0: the verb should never occur with a direct object. For the alternating category, θ takes a value between 0 and 1 inclusive. The prior probability distribution over θ in this case is a uniform $Beta(1, 1)$ distribution. We begin with the simplifying assumption that all three values of T have equal prior probability—that is, the learner assumes that any verb in the language is equally likely *a priori* to be transitive, intransitive, or alternating. In later simulations, we explore our model's behavior when this assumption is changed.

In the lower half of the model, each X is conditioned on a Bernoulli random variable e, which represents the input filter. If $e_i^{(v)} = 0$, the observation in $X_i^{(v)}$ was generated by $\theta^{(v)}$ and $T^{(v)}$, and accurately reflects the transitivity of verb v. But if $e_i^{(v)} = 1$, the observation in $X_i^{(v)}$ was generated by an erroneous parse (henceforth an "error"), meaning the learner did not have adequate grammatical knowledge to parse the sentence correctly. This observation was not generated by $\theta^{(v)}$ and $T^{(v)}$, and may not accurately reflect the transitivity of verb v, so it should be ignored for the purpose of inferring $T^{(v)}$. Each $e^{(v)}$ is conditioned on the variable ϵ, which represents the probability of an erroneous parse occurring for any sentence in the input. The model learns a single parameter value for ϵ across all verbs.

The second parameter of the input filter is δ, which represents the probability of observing a direct object when an observation was generated in error. Thus, whether a sentence contains a direct object or no direct object depends on one of two biased coins. If $e_i^{(v)} = 0$ and the observation accurately reflects the verb's transitivity properties, then one biased coin is flipped and the sentence contains a direct object with probability $\theta^{(v)}$. If $e_i^{(v)} = 1$ and the observation was generated in error, then a different biased coin is flipped

and the sentence contains a direct object with probability δ. Like ϵ, δ is a shared parameter across all verbs. We assume that both ϵ and δ have a uniform $Beta(1, 1)$ prior distribution.

3.2 Joint Inference

We use Gibbs sampling (Geman & Geman, 1984) to jointly infer the transitivity of each verb (T) and the two parameters of the input filter (ϵ and δ). In this form of sampling, we start with randomly-initialized values for ϵ and δ, and use those values to calculate the posterior probability of each transitivity category T for each verb, given the observed data and those filter parameters. We sample values for T from this posterior probability distribution. Then, we use the sampled transitivity categories to sample new values for ϵ and δ from estimates of their posterior probability distributions. This cycle is repeated over many iterations until the model converges to a stable distribution over T, ϵ and δ, which represents the optimal joint probability solution for these three variables. See the Appendix for details of the sampling procedure.

4 Simulation 1

In Simulation 1, we ask whether inferring the parameters of an input filter will allow a learner to accurately identify the transitivity categories of verbs in the speech that children hear. We tested our joint inference model on a dataset containing distributions of the 50 most frequent transitive, intransitive, and alternating verbs in corpora of child-directed speech. In order to determine whether this inference is successful, we compare our model's performance to an oracle model that already knows appropriate parameters for filtering its input, and baseline models with inappropriate filter parameters.

4.1 Data

We prepared a dataset of four corpora selected from the CHILDES Treebank (Pearl & Sprouse, 2013). This resource provides parse trees for several corpora of child-directed speech on CHILDES (MacWhinney, 2000), generated by the Charniak or Stanford parser

and hand-checked by undergraduates. The selected corpora contain 803,188 words of child-directed speech, heard by 27 children between the ages of 6 months and 5 years. See Table 1 for corpus details.

Table 1
Corpora of Child-Directed Speech

Corpus	**# Children**	**Ages**	**# Words**	**# Utterances**
Brown- Adam, Eve, & Sarah (Brown, 1973)	3	1;6-5;1	391,848	87,473
Soderstrom (Soderstrom et al., (2008))	2	0;6-1;0	90,608	24,130
Suppes (Suppes, 1974)	1	1;11-3;11	197,620	35,904
Valian (Valian, 1991)	21	1;9-2;8	123,112	25,551

Our dataset was created by extracting sentences with the 50 most frequent action verbs in these corpora that could be characterized as transitive, intransitive, or alternating. We excluded verbs with other argument-taking properties, such as obligatorily ditransitive verbs or those that frequently take clausal or verbal complements: mental state verbs (e.g. *want*), aspectual verbs (e.g. *start*), modals (e.g. *should*), auxiliaries (e.g. *have*), and light verbs (e.g. *take*).[4] We sorted the selected 50 verbs into transitive, intransitive, and alternating categories according to the English verb classes described in Levin (1993), supplemented by our own intuitions for verbs not represented in that work. These classes provide a target for learning meant to align with adult speaker intuitions, independent of the corpus data that the model learns from. The transitive and intransitive categories are conservative; any verb that could occur in a transitivity alternation was classified as alternating, regardless of the frequency or type of alternation. So, verbs like *jump* are considered alternating even though they occur infrequently in their possible transitive uses (e.g. *jump the horses over the fence*). These target categories thus set a very high bar for our model to reach.

[4]Verbs with other argument-taking properties were excluded because the current paper models only the acquisition of transitivity; including verbs with other argument-taking properties would require expanding the learner's hypothesis space to include the full set of possible argument structure classes and alternations in the language. We instead break this larger problem into smaller steps, and save for future work the question of how learners would build on transitivity knowledge to incrementally acquire this full argument structure system.

We then conducted an automated search over the Treebank trees for the total occurrences of each verb in the corpora, in all inflections, and the total occurrences with overt direct objects following the verb (right NP sisters of V). These direct object counts included basic transitive clauses, but not wh-object questions or any other sentences with object gaps. Thus, we assume a learner with the knowledge of 15- to 17-month-old English infants in previous behavioral studies: one who uses the canonical word order properties of English to identify direct objects when they occur after verbs, but does not yet know how to identify arguments in non-canonical positions. Table 2 lists the complete dataset provided to the learner: counts of the selected 50 verbs, along with their counts of overt post-verbal direct objects. For legibility we also report the percentages of direct objects with each verb, although our model learns from raw counts rather than percentages.

4.2 Results

4.2.1 Verb Transitivity Inference. Our joint inference model infers a probability distribution over transitivity categories for each verb in its dataset. These distributions are displayed in Figure 2. Black bars represent the posterior probability assigned to the transitive category, dark gray bars represent the probability assigned to the intransitive category, and light gray bars represent the probability assigned to the alternating category. The target categories for each verb are shown below the horizontal axis.

We calculated accuracy by determining which transitivity category was assigned highest probability to each verb by our model, and comparing these category assignments to the target categories for each verb. The proportion of verbs categorized correctly by the model is reported in Table 3. Overall, the model infers the correct transitivity properties for 2/3 of the verbs in our dataset. This is substantially better than chance performance: a model that randomly assigned categories to verbs would achieve 33% accuracy, because there are three possible options for each verb. Our joint inference model performs significantly better on each verb class, and nearly twice as well overall.

Table 2
Dataset: Counts and Percentage Uses with Overt Direct Objects (DO) of 50 Verbs

Verb	**Total**	**# DO**	**% DO**
Transitive Verbs			
feed	220	205	93%
fix	337	305	91%
bring	605	541	89%
throw	312	275	88%
hit	214	187	87%
buy	358	299	84%
catch	185	141	76%
hold	579	406	70%
wear	477	287	60%
Alternating Verbs			
pick	331	299	90%
drop	169	149	88%
lose	185	160	86%
close	166	141	85%
touch	183	153	84%
leave	356	297	83%
wash	195	161	83%
pull	331	268	81%
push	352	274	78%
open	342	265	77%
cut	263	198	75%
bite	191	140	73%
turn	485	350	72%
build	299	215	72%
knock	160	115	72%
read	509	350	69%

Verb	**Total**	**# DO**	**% DO**
Alternating Verbs, cont.			
break	550	347	63%
drink	366	221	60%
eat	1318	777	59%
sing	306	161	53%
blow	255	132	52%
draw	375	193	51%
move	238	112	47%
ride	281	114	41%
hang	151	53	35%
stick	192	57	29%
write	583	155	27%
fit	227	49	22%
play	1568	308	19%
stand	294	21	7%
run	228	13	6%
walk	253	11	4%
jump	197	8	4%
swim	180	7	4%
sit	859	11	1%
Intransitive Verbs			
wait	383	57	15%
work	256	11	4%
cry	275	8	3%
sleep	451	13	3%
stay	308	4	1%
fall	605	3	0%

The model achieves highest accuracy in categorizing the intransitive verbs: for all but one of these verbs, the model assigns highest probability to the intransitive category. The exception is the verb *wait*, which the model assigns highest probability under the alternating category. This is due to prevalent uses of *wait* with temporal adjuncts, as in *wait a minute*, that were indistinguishable from NP direct objects in the CHILDES Treebank parse trees. Thus, a learner who cannot differentiate these adjuncts from direct objects would infer that *wait* is an alternating rather than intransitive verb.

The model assigns 6 out of the 9 transitive verbs highest probability under the transitive category. Three transitive verbs are assigned highest probability under the

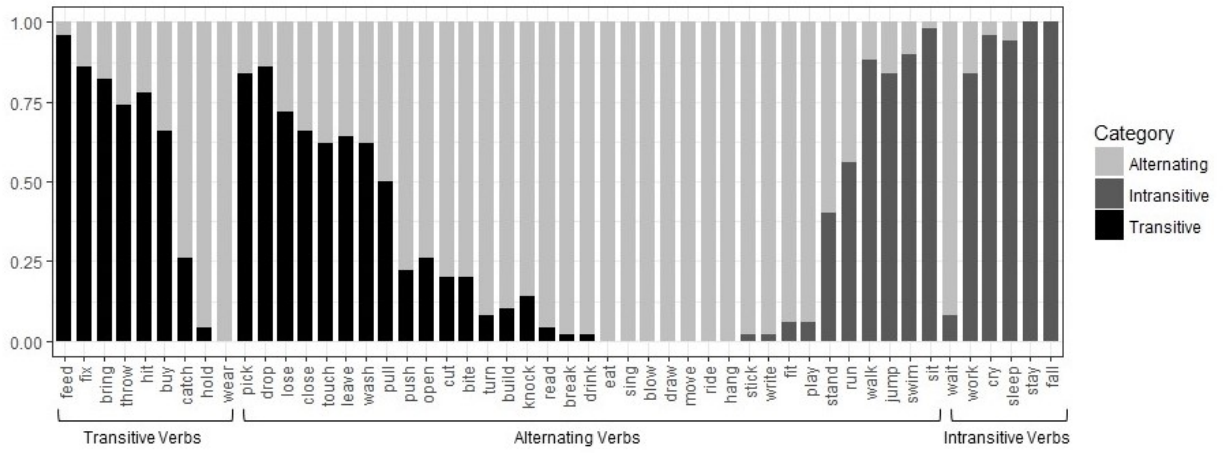

Figure 2. Posterior Distributions over Verb Categories (T), Joint Inference Model

alternating rather than the transitive category: *catch, hold,* and *wear*. This is likely because these verbs display different behavior than the other transitive verbs in the corpus. The verb *hold* occurs frequently in verb-particle constructions (e.g. *hold on*), which might be treated differently than simple verbs by learners. The verbs *catch* and *wear* appear to occur at much higher rates than other transitive verbs in non-basic clauses: *catch* occurs frequently in passives (e.g. *get caught*), and *wear* occurs frequently in wh-object questions (e.g. *what are you wearing?*). We leave for future work the question of whether children likewise mis-classify these verbs, or whether they can accommodate their different distributional behavior by using more sophisticated information than our modeled learner.

The model assigns highest probability for most of the alternating verbs to the alternating verb category. There are 13 exceptions. The verbs *pick, drop, lose, close, touch, leave*, and *wash* are assigned highest probability under the transitive category because they infrequently occur in their possible intransitive uses in child-directed speech. The verb *pull* is assigned equal probability under the transitive and alternating categories for the same reason. (This verb was not considered to be correctly assigned to the alternating category in our accuracy calculation.) The verbs *run, swim, walk, jump,* and *sit* are assigned highest probability under the intransitive category because these verbs very infrequently occur in

Table 3
Proportions of Verbs Categorized Correctly, Simulation 1

Model	**Transitive**	**Intransitive**	**Alternating**	**Total Verbs**
Joint Inference	0.67	0.83	0.63	0.66
Oracle	0.78	0.83	0.51	0.60
No-Filter Baseline	0.00	0.00	1.00	0.70
Chance	0.33	0.33	0.33	0.33

their possible transitive uses.[5] Thus, the model over-regularizes the alternating verbs that alternate infrequently, preferring the more deterministic transitive and intransitive verb categories.

4.2.2 Filter Parameter Inference. Recall that our model identifies verb transitivity categories by jointly inferring parameters for filtering its input. These parameters are ϵ, which represents the frequency of erroneous parses, and δ, which represents whether those errors are likely to cause direct objects to go missing, or to spuriously appear. Figure 3 displays the posterior probability distributions inferred by the model for ϵ and δ. In order to evaluate the model's inference of these parameters, we estimated their true value in our dataset. The proportion of transitive verbs with missing overt post-verbal direct objects in the dataset gives us an estimate of $(1 - \delta) \times \epsilon$, and the proportion of intransitive verbs with spurious direct objects (e.g. *wait a minute*) gives us an estimate of $\delta \times \epsilon$. Solving these two equations, we find that $\delta = 0.18$ and $\epsilon = 0.24$. The posterior probability distribution over δ inferred by our model has a mean of 0.25, and the probability distribution over ϵ has a mean of 0.19. Our model thus slightly over-estimates the value of δ and under-estimates the value of ϵ, but it infers values for these parameters that are close to the true values in the corpus.

[5]Note that four out of these five verbs are manner of motion verbs (*run, swim, walk, jump*), and their transitive uses do not typically involve agent-patient relations (e.g., *walk a mile*, *swim the channel*, *jump the turnstile*). Even when a causative meaning may be used, as in the case of *jump the horse*, this implies less direct causation than a typical alternating verb such as *break* or *open*. So, even though our conservative target categories treated these verbs as alternating, in some ways they behave more typically like intransitives.

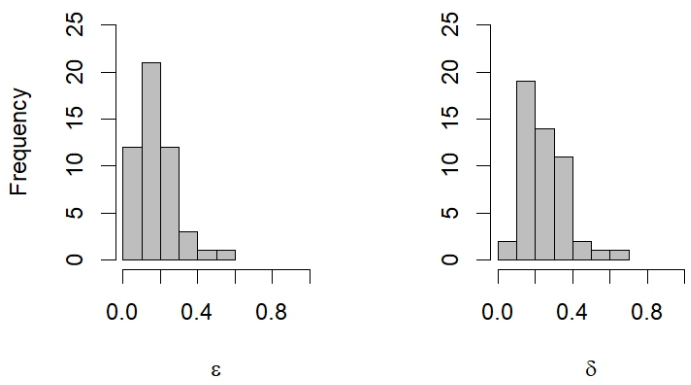

Figure 3. Posterior Distributions over ϵ and δ, Simulation 1

4.3 Model Comparisons

4.3.1 Oracle Model. The primary contribution of our model is demonstrating that a learner can filter its input without knowing anything in advance about what needs to be filtered out. Therefore, it makes sense to compare our model against an "oracle" that knows a lot about what needs to be filtered out. We instantiated an oracle model in which δ is fixed to 0.18 and ϵ to 0.24 in order to reflect their true values in our dataset, as estimated in the previous section. This oracle model thus knows the parameters for the input filter in advance: it knows how frequently erroneous parses are likely to occur, and how they will behave. By comparing our model to this oracle, we can determine whether our model's performance is impaired by having to learn these parameters.

The posterior probability distributions over verb categories inferred by the oracle model are displayed in Figure 4. The posterior probabilities inferred by the oracle are less graded than those inferred by our joint inference model; this is unsurprising, as the oracle considers only one value each for δ and ϵ instead of sampling over multiple values. But when considering which transitivity category is assigned highest probability to each verb, the two models classify most of these verbs in the same way. Our joint inference model classifies intransitive verbs identically to the oracle model, and performs almost as well with transitive

verbs: the oracle succeeds in identifying one more transitive verb, *catch*, as transitive. Our model performs better than the oracle in categorizing alternating verbs: the oracle has an even higher tendency to over-regularize the verbs that alternate infrequently. Inferring the parameters of the input filter thus results in comparable, and maybe slightly better, accuracy in categorizing verbs than knowing these parameters in advance.

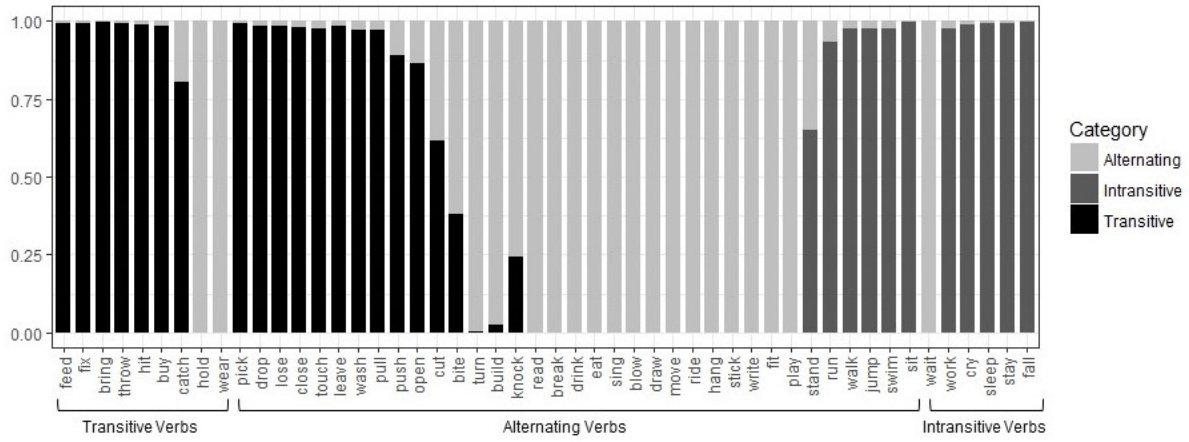

Figure 4. Posterior Distributions over Verb Categories (T), Oracle Model

4.3.2 Random Filter Parameters. If the values of the filter parameters aren't important, then it wouldn't be remarkable that our joint inference model performs comparably to the oracle model. To test whether the filter parameters actually matter, we ran 500 model simulations in which ϵ and δ were fixed to randomly-sampled values. Fig. 5 displays the model's resulting accuracy in inferring transitivity categories given each set of filter parameters, with ϵ along the x-axis and δ along the y-axis. Lighter colors denote higher percentages of verbs categorized correctly. The gray rectangle marks the range of filter parameter values that were considered highest probability by our joint inference model—specifically, these are the values within one standard deviation of the mean in the posterior probability distributions that our model inferred.

A visual scan of these plots shows that it is not trivial to infer filter parameters that will result in high accuracy across all three transitivity categories. Higher values of ϵ yield higher accuracy on categorizing transitive and intransitive verbs, but lower accuracy for

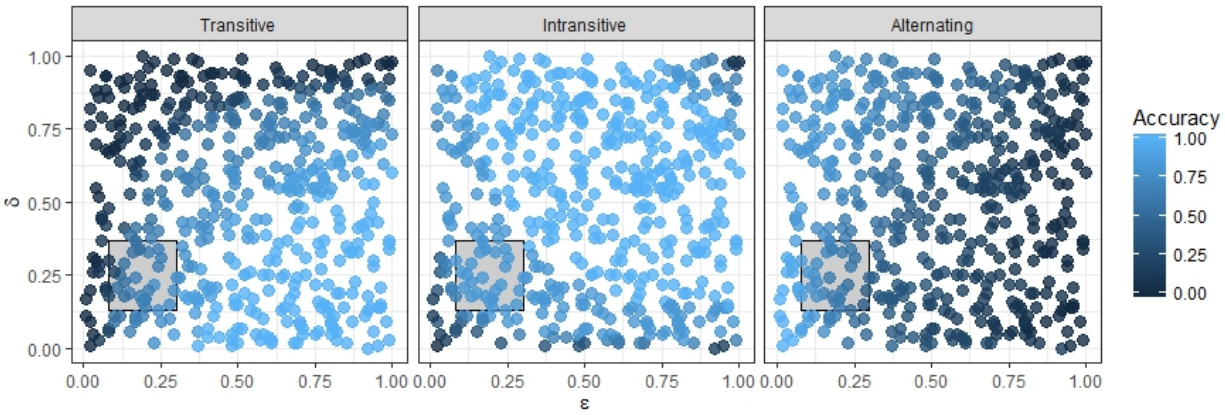

Figure 5. Accuracy (% Verbs Categorized Correctly) by Varying Values of ϵ and δ

alternating verbs. This is because the learner assumes there is more error in its transitive and intransitive verb observations, and lowers the threshold for assigning verbs to those categories. The learner thus assigns more verbs in its dataset to the transitive and intransitive categories rather than the alternating category. On the other hand, higher values of δ yield lower accuracy for transitive verbs, but higher accuracy for intransitive verbs. With higher values of δ, the learner assumes that more of its errorful sentence observations contain mistaken direct objects, rather than missing direct objects. The learner therefore expects more error in its intransitive verb observations because there should be more intransitive verbs appearing with spurious direct objects. This lowers the threshold for assigning verbs to the intransitive class, resulting in higher accuracy for intransitives. Conversely, the learner expects less error in its transitive verb observations because there should be fewer transitive verbs appearing with missing direct objects. This raises the threshold for assigning verbs to the transitive class, resulting in lower accuracy for transitives.

Thus, successfully categorizing verbs in all three transitivity classes requires inferring filter parameters that fall within a somewhat narrow range. Our model performs comparably to the best-case oracle model not merely because it infers an input filter, but because it infers the best parameters for such a filter given our dataset. Note that our model is not actually optimizing for the accuracy values plotted in the graph in Fig. 5, because it is not trained on our target classifications for verb transitivity. Instead, the model is optimizing for probability:

it is searching for the best joint-probability solution for verb transitivity categories and filter parameters to explain the distributions in its data. The fact that our model performs well with respect to our target verb classifications means that the parameter values that have high probability under our model also result in good accuracy across all three verb classes.

4.3.3 No-Filter Baseline. Our model accurately categorizes verbs across transitivity categories by inferring appropriate parameters for a filter on its input, and the model comparisons above show that the values of these filter parameters are important. Models with grossly inappropriate filter parameters might have better accuracy on some verb classes, but do not perform as well across all three transitivity categories. A special case of these models would be those where ϵ equals exactly zero, representing zero probability of parsing errors: this produces models that do not have an input filter at all. Comparing against a no-filter baseline tells us how much having a filter matters in identifying verb transitivity.

As values of exactly zero were never randomly sampled in the simulations reported in Fig. 5, we conducted an additional simulation setting ϵ to zero. The value of δ in this case does not matter, because it is never used. Because every verb in our dataset occurs some but not all of the time with overt post-verbal direct objects, and this no-filter model assumes there are no parsing errors to filter out, it assigns every verb to the alternating category. It thus categorizes 100% of the alternating verbs correctly, achieving 70% overall accuracy because alternating verbs make up 70% of our dataset. However, this accuracy comes at the cost of failing to categorize any verbs as transitive or intransitive. Our joint inference model performs substantially better in this regard, categorizing the majority of transitive and intransitive verbs correctly. This demonstrates that an input filter is important for differentiating alternating from non-alternating verbs.

4.3.4 Threshold Comparisons. By inferring how frequently parsing errors occur in its sentence observations and the behavior of those errors, our model is essentially inferring where to put thresholds for classifying verbs as transitive or intransitive based on

rates of observed direct objects. Another way of evaluating our model's performance is to compare it against a simple threshold model, which classifies verbs as transitive if their percentage occurrence with overt direct objects falls above a certain threshold, and as intransitive if their percentage occurrence with overt direct objects falls below a certain threshold. There are several differences between this type of threshold model and our model. Instead of setting hard thresholds that delineate each of these categories, our model uses soft thresholds that take into account how much data it has available for any particular verb. And the primary advance in our model is that these soft thresholds are learned: the model does not need to know the true distributions of transitive and intransitive verbs in advance. If our model performs comparably to a model that knows the best thresholds for classifying its data, this will give us another indication that it is learning successfully.

To create these comparisons, we hand-fit the thresholds for classifying verbs by percentage overt direct objects to maximize accuracy on the model's dataset. Table 4 reports the accuracy of the best-performing threshold models, compared to our joint inference model. The thresholds that yielded the best performance overall were 87% and 4%: this model classifies verbs as transitive if they occur with direct objects above 87% of the time, and verbs as intransitive if they occur with direct objects less than 4% of the time. This model was able to achieve 80% accuracy overall. However, its performance on classifying transitive and intransitive verbs was lower than for our joint inference model. Our second threshold comparison thus aimed to maximize overall accuracy without performing lower than our joint inference model on these two verb classes. Thresholds of 83% and 5% allowed the model to achieve 72% overall accuracy, while achieving the same accuracy as our joint inference model on transitive and intransitive verbs. Finally, our third threshold comparison attempted to maximize overall accuracy while achieving *higher* accuracy than our joint inference model on transitive and intransitive verbs. The best thresholds for this model were 76% and 15%. This threshold model's higher performance on transitive and intransitive verbs led to lower accuracy on alternating verbs, and it only achieved 64% accuracy overall.

Table 4
Proportions of Verbs Categorized Correctly: Best-Performing Threshold Models

Model	**Transitive**	**Intransitive**	**Alternating**	**Total Verbs**
Joint Inference	0.67	0.83	0.63	0.66
Thresholds of 87%, 4%	0.56	0.66	0.89	0.80
Thresholds of 83%, 5%	0.67	0.83	0.71	0.72
Thresholds of 76%, 15%	0.78	1.00	0.54	0.64

Although our joint model is not explicitly learning thresholds, we can use the filter model parameters that our model inferred to estimate the soft thresholds it is effectively using. Because ϵ is the inferred rate of error and δ is the inferred proportion of error that has direct objects, $\epsilon \times (1 - \delta)$ gives an estimate of the rate of missing direct objects for transitive verbs. Therefore, $1 - \epsilon \times (1 - \delta)$ can be interpreted as a threshold of direct object rates above which verbs are more likely classified as transitive. Conversely, $\epsilon \times \delta$ estimates the rate of spurious direct objects for intransitive verbs, and thus provides an estimate for a threshold below which verbs are more likely classified as intransitive. When we estimate thresholds based on the means of the distributions over ϵ and δ that our model inferred (0.19 and 0.25), we obtain estimated thresholds of 85% and 5%. These are very close to the thresholds that yielded the best performance in our threshold models.

In summary, these comparisons show that it is possible for a simple threshold model to achieve higher overall accuracy than our joint inference model, if it is allowed to use thresholds that are hand-fit to maximize performance on this dataset. However, it is not trivial to find hard thresholds that will ensure high performance across all three verb classes. In particular, the best-performing threshold models may have exceeded the overall accuracy of our joint inference model, but they never exceeded our model's accuracy on both transitive and intransitive verbs without reducing overall accuracy. This shows us that the soft thresholds that our model is essentially learning are appropriate to its dataset: our model performs just as well as the best-performing threshold models on identifying these deterministic verb categories. And this is true even though our model is not optimizing for

accuracy. Unlike the threshold models, our model does not have access to the target classifications for verb transitivity in its dataset, and cannot use those classifications to identify its thresholds. Instead, our model learns where to put these soft thresholds by finding the best joint probability solution for verb transitivity categories and the parameters for error in its dataset.

4.4 Discussion

Our model accurately categorizes 2/3 of the most frequent transitive, intransitive, and alternating verbs in child-directed speech on the basis of their distributions with and without direct objects, by learning to filter out sentences that were likely mis-parsed. This enables the learner to avoid drawing faulty inferences about verb transitivity from non-basic clause types that may be mistaken for intransitive clauses. Our model performs comparably to an oracle model that knows in advance the best parameters for a filter given its dataset, and better than many models with inappropriate filter parameters. It performs substantially better in categorizing transitive and intransitive verbs than a baseline model that lacks an input filter altogether, and performs twice as well overall as would be expected by chance. It also performs just as well on categorizing transitive and intransitive verbs as the best-performing threshold models, which categorize verbs using thresholds of direct object rates that are hand-fit to the dataset. These results demonstrate that an input filter both matters for verb transitivity learning, and can be learned.

The model makes two types of mistakes in inferring verb categories. First, it is unable to correctly categorize some transitive and intransitive verbs that behave differently than other verbs in their category, such as *catch, hold, wear*, and *wait*. Further investigation is necessary to determine whether these verbs pose difficulties for child learners as well. A second type of mistake is over-regularizing alternating verbs that alternate infrequently: the model prefers to assign these verbs to the transitive and intransitive categories. This is an example of a learner preferring a more deterministic analysis for probabilistic input, a

tendency also found in child learners in artificial language studies (Hudson Kam & Newport, 2009). The error-filtering mechanism we present here could thus potentially provide a way to model other forms of over-regularization in learning.

There are three factors that contribute to our model's ability to regularize its input. First, our learner only needs to infer two parameters for its input filter: it makes the simple assumption that there is a single value for ϵ and δ shared across all verbs, rather than having to infer separate values for these parameters on a verb-by-verb basis. This allows the learner to use distributions of direct objects across verbs to inform its estimates of how much error is present in its sentence representations, and what that error looks like. If instead the learner expected a different ϵ and δ for each verb, it would be difficult for the learner to tell whether a particular rate of direct objects observed for a verb is due to a particular rate of transitivity alternation (θ) or due to a particular type of error that occurs only with that verb.

Intuitively, the expectation of a single shared value for these filter parameters corresponds to the expectation that the noise process generating the error in the learner's sentence representations reflects some properties that are independent of the particular verbs in those sentences. We believe that this expectation is not only a helpful simplification, but also a realistic one. While our learner has no commitment to what this noise process is, in reality it reflects the contribution of a variety of grammatical operations that the learner has mis-parsed. These operations are due to independent properties of the grammar, and apply to entire classes of verbs, not on a verb-by-verb basis. A more sophisticated learner might identify that there are several noise processes at work, corresponding to these different grammatical properties, and use distributions of direct objects across verbs along with other surface features of these sentences to infer a different ϵ and δ for each of these properties.

Additionally, the learner's inference of its input filter is successful because it encounters a wide variety of verb behavior in its data. Some verbs appear more deterministic than others: they alternate less frequently, instead show a stronger preference for solely transitive or intransitive frames. Just as we used the true transitive and intransitive verbs in the

dataset to arrive at our estimates of the true values for ϵ and δ, our learner can anchor its estimates of these parameters by using the distributions of direct objects with the more deterministic verbs it observes— those that it thinks are more likely to be transitive or intransitive. If instead all verbs alternated at exactly the same rate, the learner would have difficulty knowing whether all verbs have exactly the same transitivity properties, or whether there is additional error present. This raises the question of whether all languages have enough variety in verb distributions to enable successful learning by this filtering mechanism. Answering this question would require testing this model with cross-linguistic corpora of child-directed speech, a future direction that we discuss more in the General Discussion.

Finally, our learner's ability to successfully regularize depends on having deterministic categories in its hypothesis space: it expects that some verbs will only occur in transitive or intransitive frames, and makes the simplifying assumption that these verbs are equally likely *a priori* as verbs that can alternate. However, we might ask how realistic it is for a learner to have this assumption, as in reality these categories will occur in different proportions in the target language. Will a learner perform just as well if it expects transitive, intransitive, and alternating verbs to occur with different frequency? We can answer this question by examining the model's performance when it has different prior beliefs about the probability of these verb classes. If there is no difference in performance, then it suffices to merely have transitive or intransitive categories in the learner's hypothesis space, regardless of how they are weighted. But if there is a difference in performance, this would show that the model's prior beliefs about the relative probabilities of transitivity classes matter for its ability to identify these classes in its input.

5 Simulation 2

In Simulation 2, we ask whether our model will still accurately identify the transitivity categories of verbs in child-directed speech if it does not expect transitive, intransitive, and alternating verbs to be equally likely *a priori*. Instead of setting a uniform prior over

transitivity categories ($P(T^{(v)})$ in Equation 7), we biased the model's prior in favor of alternating verbs. In Simulation 2a, we set the model's prior to match the actual frequencies of verb transitivity categories in its input: we set a prior probability of 0.70 for alternating verbs, 0.18 for transitive verbs, and 0.12 for intransitive verbs, to match the proportion of the target verb categories in our dataset. This allows us to determine whether our learner's verb transitivity inference is affected if it expects to find verb categories in the same proportions as they will actually occur in its input. In Simulation 2b, we skewed the model's prior even more heavily in favor of the alternating category: we set a prior probability of 0.90 for alternating verbs and 0.05 each for transitive and intransitive verbs. By giving the alternating category substantially greater prior probability than the two deterministic verb categories, we can determine whether simply having transitive and intransitive categories in the learner's hypothesis space, in any proportion, is sufficient for identifying them in its input.

5.1 Data

We tested our skewed-prior models on the same dataset of transitive, intransitive, and alternating verbs in child-directed speech that we prepared for Simulation 1.

5.2 Results

5.2.1 Verb Transitivity Inference. Fig. 6 displays the posterior probability distribution over transitivity categories that our model inferred for each verb in Simulation 2a, when it expected 70% alternating verbs. Fig. 7 displays the distribution over transitivity categories inferred in Simulation 2b, when the model expected 90% alternating verbs. Table 5 reports the proportion of verbs categorized correctly in each transitivity category, compared to our original joint inference model in Simulation 1.

In Simulation 2a, the inferred distribution over transitivity categories is very similar to the distribution inferred by our original model in Simulation 1. This model assigns highest probability under the transitive category to the same 6 out of 9 transitive verbs as our

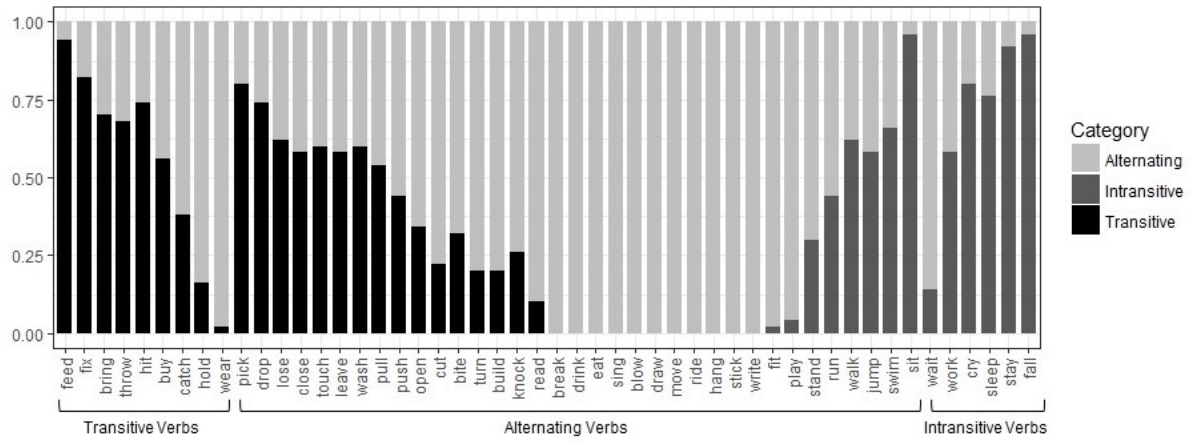

Figure 6. Posterior Distributions over Verb Categories (T), Simulation 2a

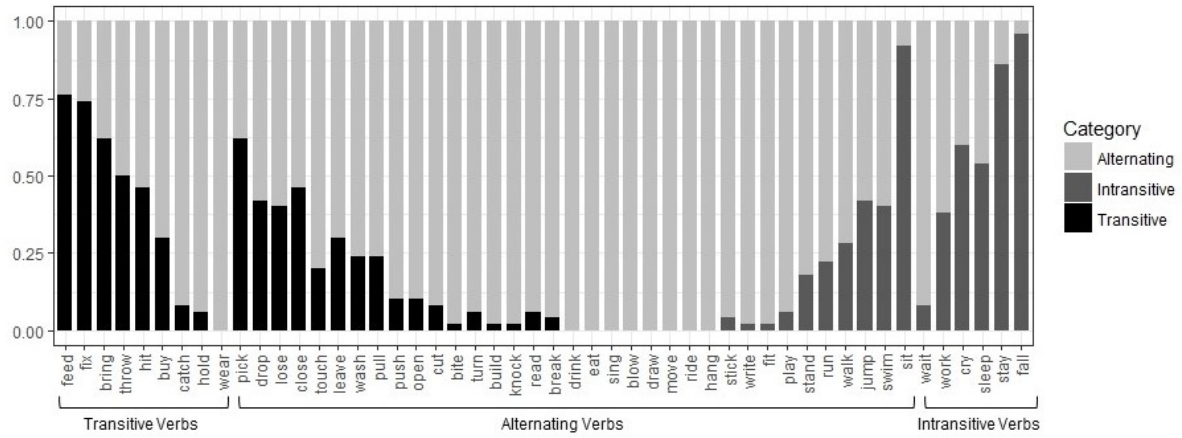

Figure 7. Posterior Distributions over Verb Categories (T), Simulation 2b

original model, and it assigns highest probability under the intransitive category to the same 5 out of 6 intransitive verbs. The model also assigns highest probability under the alternating category to 23 alternating verbs, and considers the remaining 12 to be either transitive or intransitive, over-regularizing at nearly the same rate as our original model. Thus, skewing the model's prior to expect alternating verbs 70% of the time resulted in very little difference in verb categorization accuracy compared to our original model.

In Simulation 2b, when we skewed the model's prior to expect alternating verbs 90% of the time, the model inferred a different distribution over transitivity categories. There are two general trends to observe in these data. First, even though this learner was heavily

Table 5
Proportions of Verbs Categorized Correctly, Simulations 1 and 2

Model	**Transitive**	**Intransitive**	**Alternating**	**Total Verbs**
Simulation 1	0.67	0.83	0.63	0.66
Simulation 2a	0.67	0.83	0.66	0.68
Simulation 2b	0.33	0.67	0.94	0.80

biased against the transitive and intransitive categories, there are still several verbs that it assigns high probability under these categories. To some extent, the model was able to overcome its biased prior and identify some deterministic verbs in its input.

On the other hand, there are fewer verbs that this model assigns highest probability under the transitive and intransitive categories, and more verbs that it assigns highest probability under the alternating category. This results in higher accuracy for alternating verbs: this model only over-regularizes one of these verbs (*pick*) as transitive, and one of these verbs (*sit*) as intransitive. Because alternating verbs are most frequent in the model's data, the model's higher accuracy on alternating verbs leads to higher total accuracy as well. But the model achieves lower accuracy for the transitive and intransitive categories. The model assigns highest probability to the transitive category for only 3 of the 9 transitive verbs, and it assigns highest probability to the intransitive category for only 4 of the 6 intransitive verbs. Of the target transitive verbs, the model now considers *throw, hit*, and *buy* to be alternating, along with *catch, hold*, and *wear*. Of the intransitive verbs, the model now considers *work* to be alternating along with *wait*. The model still performs better than chance in categorizing intransitive and alternating verbs, but it is no different from chance in categorizing transitive verbs.

In summary, we found comparable performance to our original model when we skewed the model's prior to expect transitive, intransitive, and alternating categories in the same proportions as they actually occur in the input. However, when we biased the model more strongly towards the alternating category, it identified transitive and intransitive verbs at a

much lower rate. The model's rate of regularization was not affected by its bias against deterministic categories in Simulation 2a, but was affected by its stronger bias in Simulation 2b.

5.2.2 Filter Parameter Inference. Figs. 8 and 9 display the posterior probability distributions over ϵ and δ inferred by the skewed-prior models. Although the shapes of these distributions are different, they are centered around similar values as those inferred by our original model in Simulation 1. The mean of the distribution over ϵ is 0.22 in Simulation 2a and 0.19 in Simulation 2b, compared to 0.19 for our original model. The mean of the distribution over δ is 0.23 in Simulation 2a and 0.21 in Simulation 2b, compared to 0.25 for our original model. Just as for our original model, these values are close to the estimated true values of $\epsilon = 0.24$ and $\delta = 0.18$ in the model's dataset, as calculated for Simulation 1.

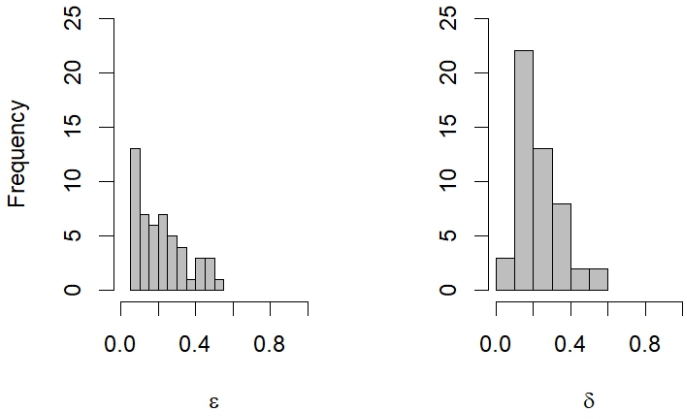

Figure 8. Posterior Distributions over ϵ and δ, Simulation 2a

Thus, changing the learner's prior beliefs about how transitivity categories distribute in its input did not substantially affect its inference about the parameters of its input filter: it still inferred appropriate values for the frequency and behavior of error in its data. This might be because the learner is anchoring that inference on the distributions of the verbs that it considers to be transitive and intransitive with the highest probability. Because both

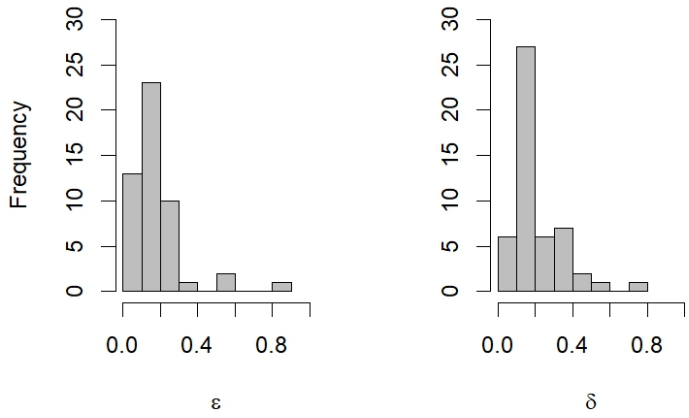

Figure 9. Posterior Distributions over ϵ and δ, Simulation 2b

models in Simulation 2 did identify some transitive and intransitive verbs, and those verbs are a subset of the verbs that our original model categorized as transitive and intransitive with highest probability, it is not so surprising that all three models found similar parameters for their input filters. Moreover, inferring these parameters is what allowed the model in Simulation 2b to still categorize some verbs as transitive and intransitive, despite its strong bias against those categories. Without a filter, the model would perform identically to the no-filter baseline in Simulation 1, and categorize all verbs as alternating.

5.3 Model Comparison: Random Prior Parameters

We found different results for the model's verb transitivity inference depending on how much we biased its prior against transitive and intransitive verbs. This raises the question: under what circumstances does the model's prior substantially affect its ability to identify verb transitivity, and under what circumstances does it not matter? That is, how much bias against deterministic verb categories can our learner accommodate and still accurately identify those categories in its input?

To answer this question, we ran 500 model simulations in which the model's prior probabilities over transitive, intransitive, and alternating categories were fixed to randomly

sampled values that summed to 1. Because the models in Simulations 1 and 2 inferred similar values for ϵ and δ, for ease of computation we set these filter parameters to the mean values of $\epsilon = 0.20$ and $\delta = 0.23$ that were inferred in those previous simulations. Fig. 10 plots the learner's accuracy in categorizing transitive, intransitive, and alternating verbs as its prior becomes more skewed towards the alternating category. The x-axis displays varying values of the model's prior on alternating verbs, and the y-axis displays the average percentage of verbs in each class categorized correctly at each of those values. A curve of best fit is plotted using a running LOESS regression (local nonparametric regression; Cleveland & Devlin, 1988).

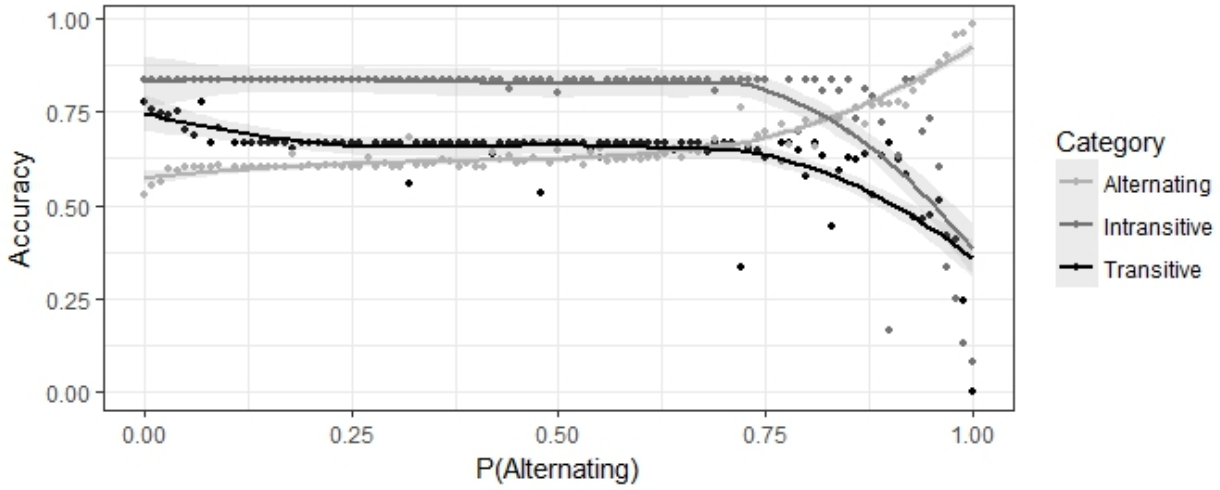

Figure 10. Accuracy (% Verbs Categorized Correctly) by Prior Probability on Alternating Verbs

This plot shows that the learner's accuracy in verb categorization remains steady across a large range of prior parameter values. When its prior probability on alternating verbs is less than approximately 0.75, the learner's performance is fairly consistent: it correctly categorizes on average 6/9 transitive verbs, 5/6 intransitive verbs, and 22/35 alternating verbs. Performance only begins to vary when its prior probability on alternating verbs is pushed above 0.75. Above this value, its accuracy on categorizing transitive and intransitive verbs declines and its accuracy on alternating verbs increases, as it categorizes fewer verbs as transitive and intransitive. Thus, it appears that there is a large range of bias

towards or against deterministic verb categories that our learner can accommodate without affecting its ability to identify those verbs in its input. It only begins to lose that ability when its bias against deterministic categories becomes extreme.

5.4 Discussion

While Simulation 1 shows that an appropriate input filter is important for learning verb transitivity, Simulation 2 shows that learning is also affected to some extent by the learner's prior beliefs about the relative frequency of transitivity categories in its input. Skewing the model's prior to expect verb transitivity categories in the same proportions that it would actually encounter in its input did not affect its performance; its accuracy in categorizing transitive, intransitive, and alternating verbs was nearly identical to our original model. However, skewing the model's prior more extremely in favor of alternating verbs resulted in different performance. With a heavy bias against deterministic categories, the model over-regularized alternating verbs much less, leading to higher accuracy on that verb class and higher accuracy overall. But the model was also less successful at identifying the target transitive and intransitive verbs, and did not perform above chance levels in categorizing transitive verbs.

This behavior reveals two properties of our learner. First, it did not matter whether the learner expected transitive, intransitive and alternating verbs to be equally likely *a priori*, or whether it expected them to occur in the same proportions as they actually do occur in child-directed English. In fact, it appears that our model's performance would be very similar across a large range of prior parameters. It is desirable that our learner can succeed at identifying verb transitivity without prior expectations that match the proportions of transitivity categories in the input— this will allow the learners to be somewhat flexible in learning different target languages, even if transitivity categories distribute differently in those languages compared to English or compared to the learners' own priors. However, there is a point where the learner's prior does exert an influence on its

verb categorization. When it was extremely biased to expect alternating verbs, our learner was not able to successfully categorize transitive verbs. This means that merely having deterministic categories in the learner's hypothesis space, in any proportion, does not suffice for accurately identifying those categories in the learner's input. A learner must give those categories sufficient prior weight in order to find them.

Second, even a learner strongly biased in favor of alternating verbs was able to infer appropriate parameters for filtering sentences that were likely mis-parsed. This allowed it to identify at least some of the transitive and intransitive verbs in its input, and to avoid drawing the mistaken inference that all verbs are alternating. This filtering was less effective for a learner with an extreme bias against the transitive and intransitive categories: its bias hampered its ability to detect the signals of these deterministic categories in the data that it let through its filter. However, the fact that the learner inferred appropriate filter parameters even in this case points towards a promising direction for future research. A more sophisticated learner might incrementally update its prior over transitivity categories given more evidence about their distribution in its input, inferring the parameters of that distribution in a hierarchical model. In this case, the learner's correct initial estimates of its input filter parameters could be very helpful in identifying the right distribution over transitivity categories. Thus, even if a learner's prior beliefs about transitivity are grossly inaccurate, inferring an input filter might allow it to appropriately adjust those beliefs as it learns more about its language.

6 General Discussion

Learning in any domain depends on how the data for learning are represented. To the degree that representations of the input change over development, either due to learning or maturation, this will have an impact on how learners form categories and generalize from their data. We examine this phenomenon in the domain of language acquisition, focusing on an apparent paradox concerning the input to argument structure learning.

Learners use verbs' distributions in transitive and intransitive clauses to draw inferences about their argument-taking properties and meanings. Non-basic clauses interfere with these inferences because young learners might not recognize when arguments of the clause have been displaced from their canonical positions, and therefore might not represent clause transitivity when it is present. We have followed a proposal that children need to filter non-basic clauses out of the data they use for verb learning (Lidz & Gleitman, 2004a, 2004b; Pinker, 1984, 1989), but this creates an apparent paradox. Identifying the structure of non-basic clauses—in which transformations have applied to displace clause arguments—would seem to depend on knowing some of the core argument structure properties of the language, and yet learners need to filter non-basic clauses in order to bootstrap their learning of those very properties. Empirical findings indicate that this paradox is not merely hypothetical, but is faced by learners prior to their second birthdays. Experimental work suggests that English-learning 1-year-olds begin acquiring basic argument structure slightly before they learn to identify displaced arguments in common non-basic clause types (Gagliardi et al., 2016; Jin & Fisher, 2014; Lidz et al., 2017; Perkins & Lidz, 2020; Perkins, 2019; Seidl et al., 2003).

We offer a new solution to this paradox, which does not require the learner to detect any direct or indirect signals to non-basicness (Pinker, 1984, 1989; Gleitman, 1990). We instantiate a learner that considers the possibility that it occasionally parses sentences erroneously. The learner infers how to filter out errors from the data it uses for verb learning, without knowing where those errors came from. It observes only verbs' distributions with and without direct objects, and does not track any additional syntactic or non-syntactic cues that might correlate with non-basicness—to this learner, a wh-object question is indistinguishable from an intransitive clause. Nonetheless, our model successfully infers appropriate parameters for filtering its input in order to identify the transitivity of the majority of frequent verbs in child-directed speech. We therefore demonstrate that it is possible for a learner to filter non-basic clauses for verb learning, without knowing which

clauses are non-basic and without needing to infer what the features of non-basic clauses are. This provides an account for how the first attested steps of verb argument structure learning in infancy can take place even as non-basic clause acquisition is still developing.

More broadly, by introducing a mechanism for a learner to filter erroneous parses of its input, our model helps answer what has remained an open question in bootstrapping and verb learning: how learners manage to avoid drawing faulty inferences about grammar and meaning, at stages of development when they lack the linguistic knowledge to arrive at veridical syntactic representations of sentences they hear. This ability has been traditionally assumed by theories of both syntactic and semantic bootstrapping (Lidz & Gleitman, 2004a, 2004b; Gleitman, 1990; Pinker, 1984, 1989), and has been presupposed by previous computational models of verb learning (Alishahi & Stevenson, 2008; Barak et al., 2014; Parisien & Stevenson, 2010; Perfors et al., 2010). These previous models assume that learners can veridically represent the arguments in a clause, and use those syntactic percepts to identify verbs' core argument-taking properties and their ability to productively generalize across different argument structure alternations. Our model addresses the question of how this process begins. We propose that a learner equipped with a filtering mechanism can still identify a verb's basic argument structure, even before that learner can reliably identify all of the arguments in sentences she hears.

Our case study focuses on only one argument structure property— transitivity— but one that is arguably at the core of early grammar learning. The categories 'subject' and 'object' form the core arguments of the clause, providing the skeleton for infants' earliest clause structure representations. Furthermore, transitivity is robustly correlated with clause meaning cross-linguistically, making it a particularly useful cue for early verb learning (Fisher et al., 2010; Gleitman, 1990; Hopper & Thompson, 1980; Lidz & Gleitman, 2004a; Naigles, 1990). Although other argument structures and alternations, such as datives, have received considerable attention in prior literature (Baker, 1979; Barak et al., 2014; Parisien & Stevenson, 2010; Perfors et al., 2010; Pinker, 1989), many of these alternations involve

more language-specific and idiosyncratic form-meaning relations. These alternations are thus less central to the core problem that syntactic and semantic bootstrapping proposed to solve: how to initially break into a grammatical system whose abstract representations can be realized as many different surface forms. At the onset of learning, principled correlations between syntactic and conceptual categories, such as those that exist between transitivity and causative meanings, might provide the learner with the foothold needed to bootstrap into this system.

Our model diverges from previous computational models of bootstrapping (Abend et al., 2017; Kwiatkowski et al., 2012; Maurits et al., 2009) by learning from a very limited type of data. Our learner identifies verb transitivity only by using rates of overt direct objects, and does not have access to any additional syntactic or non-syntactic features of the sentences or the discourse environment. By limiting our learner's data in this way, we do not imply that real-life learning proceeds only from this type of distributional information. On the contrary, it is likely that children make simultaneous use of a much fuller set of information in inferring a grammar, including conceptual representations of the extra-linguistic contexts of the sentences they hear. But by investigating how much can be learned solely from verbs' syntactic distributions, we are testing the viability of the proposal that infants can use syntactic information to draw helpful generalizations even if they do not know which event in the world a particular sentence describes (Gleitman, 1990).

This issue has not been fully examined in prior bootstrapping models, which assume that learners begin by accessing the exact meaning (or set of possible exact meanings) of a sentence, represented under a structure that is homomorphic with the syntactic structure (Abend et al., 2017; Kwiatkowski et al., 2012; Maurits et al., 2009). Given access to this full conceptual representation, or instead to the full syntactic representation of a sentence, these models show that it is simple to learn how to convert from one representation to the other. This is because the learner's meaning representation is in a form that encodes all and only the predicate-argument relations in the syntactic representation of the sentence, and there is

an assumption built into the learner that those two representations will mirror each other. The bootstrapping task thus reduces to the problem of identifying which lexical items express which predicates and arguments in the learner's conceptual structure. Given this information, the learner can infer the syntactic representation of a sentence by reading off of its structured conceptual representation, and vice versa.

But bootstrapping is not so simple if learners only have access to approximations of these representations, or if conceptual structures encode more relations than those expressed in the sentence's argument structure. Even if children can perceive events and event relations in the world in the same way as adults do, it is not straightforward to identify which event relations a sentence expresses solely from its context of use (Gleitman, 1990). And when we consider the wide range of syntactic relations that might be instantiated in a particular sentence, including the various non-local dependencies found in non-basic clauses, it seems even less straightforward for the child's non-linguistic perception of the world to yield a meaning in a form that is homomorphic with the syntax of that sentence. Here, we ask whether learning can still succeed in cases where a child might not have access to conceptual and syntactic representations that mirror each other in their structure. If either of these representations is approximate or incomplete, then children must use whatever partial information might be useful in one domain— syntax or meaning— as probabilistic evidence for drawing inferences about the other domain. We show how learners might accommodate error in their syntactic percepts, such that those percepts are still useful as evidence for drawing further generalizations about their language.

Crucially, an input filtering mechanism like the one we propose can flexibly adapt over the course of a learner's development. As a learner gains more knowledge of the grammar of her language, her syntactic percepts will change: she will be learning from more complete and more accurate parses of the sentences she hears. This means that the error in her syntactic percepts will also reduce over time, and she will not need to filter as much of her data for learning. In our case, our model is learning from data that reflects the parses of an

immature learner at a particular stage of development: one who cannot identify objects when they are realized in non-canonical positions, and who mistakes certain NP adjuncts for arguments. These data do not veridically reflect the distributions of verbs with direct objects in the actual input to the learner. Thus, the learner is not inferring filter parameters to fit its actual input— instead, the learner is inferring filter parameters to fit its erroneous representations of that input. A more mature learner who has learned to identify argument displacement in English will have access to a different dataset, one that has a lower rate of error. This more mature learner would identify different parameters for filtering its data in order to learn more about its grammar.

Our model is merely a starting point, beginning with one corner of English argument structure. But having presented a proof of concept that our filtering solution is possible, we can ask how far it could generalize. In future work, we aim to test whether this model could be extended to languages with freer word order or rampant argument-drop. These linguistic properties may make it difficult for learners to identify clause transitivity even in simple, active, declarative clauses. For example, the relatively free word order of Japanese compared to English means that word order is less helpful for identifying subjects and objects in a clause, and learners must use language-specific case morphology instead; furthermore, the ability of Japanese speakers to freely drop the subject and/or object of a clause if it is salient in the discourse means that a learner must use discourse cues to recognize when silent arguments are present. For these reasons, Japanese and other languages with some of these properties, like Mandarin and Korean, are potentially problematic for syntactic bootstrapping strategies that rely on learners accurately identifying transitive verbs (Lee & Naigles, 2005, 2008), but see Fisher, Jin, and Scott (2019) and Suzuki and Kobayashi (2017) for evidence that learners do nonetheless succeed. If our model can learn appropriate parameters for filtering out the relatively higher rate of potentially misleading data in languages like Japanese, this may help clarify how syntactic bootstrapping is possible in these languages.

An additional question for future work is how children learn to identify the structure of non-basic clauses in their language. How do learners identify which transformations are present in sentences that may have initially been parsed in error? Following Gagliardi et al. (2016) and Stromswold (1995), Perkins (2019) argues that verb transitivity may be an important first step: if a learner expects a particular argument for a verb and encounters sentences where that argument does not appear in its canonical position, the learner may be compelled to examine those sentences to determine the cause of the missing argument. Thus, a strategy of identifying sentences that were likely parsed in error may help learners not only filter their input for learning verb transitivity, but also eventually learn how the target language realizes various grammatical transformations.

More broadly, we might ask whether this filtering mechanism could generalize beyond verb transitivity learning, to other cases in language acquisition where learners must ignore misleading data in order to draw correct inferences about their language. For example, prior work has proposed that some form of input filtering is helpful in identifying vowel categories (Adriaans & Swingley, 2012), and in drawing the right generalization about the constraints on the antecedent of anaphoric *one* in English (Pearl & Lidz, 2009). Filtering may also provide a mechanism for understanding why young learners tend more strongly than adults to regularize probabilistic input in artificial language studies (Hudson Kam & Newport, 2009), and how learners can acquire correct generalizations about their first language from noisy input by second-language speakers (Singleton & Newport, 2004; Schneider, Perkins, & Feldman, 2020). When our learner expects that error might be masking regularities in its data, filtering allows it to identify those regularities, and even to over-regularize in some cases. Thus, a combination of determinism in children's hypothesis spaces, along with the expectation of error in their input representations, may help explain when children draw deterministic generalizations about their language and how they draw the right ones.

Finally, the filtering mechanism we propose offers a new perspective on the use of data in learning. Typically-developing children acquire a language on the basis of only a few years'

worth of linguistic input— far exceeding the ability of our most advanced language processing technologies, and using only a fraction of the data that is necessary to train those systems. Despite the received wisdom that more data is always better, our case study suggests that children's success may be in part due to their ability to be strategic about what data to learn from. By suggesting an advantage to learning from smaller data, this filtering mechanism is similar in some ways to Newport (1990)'s "Less is More" hypothesis, under which young children's language learning is facilitated by extralinguistic cognitive limitations that restrict the amount of data they can process. But our model's filter differs in an crucial way: instead of being a by-product of external processing limitations, this filter is an integral part of the learning mechanism, arising from the learner's assumption of a noisy relationship between its data and the hypotheses it is evaluating. Under our approach, learners jointly infer the regularities underlying a particular phenomenon in their input, and what data to use in order to best identify those regularities. This type of input filtering is with respect to a specific learning goal— a child attempting to acquire a different phenomenon might filter her input in an entirely different way— and therefore provides more flexibility than an approach that imposes a hard constraint on the amount of data a learner can access. This flexibility invites further investigation into how broadly this filtering mechanism might generalize beyond language learning: it is possible that we might find strategic input filtering in learning in many other domains in which learners must generalize from noisy or unreliable data. Understanding when learners choose to learn from their input, and when they choose not to learn, may help illuminate why learning in these cases can be so remarkably successful.

7 Acknowledgments

This work was supported by the National Science Foundation (#BCS-1551629, Doctoral Dissertation Improvement grant #BCS-1827709, and NRT award #DGE-1449815). We thank Lillianna Righter, Alexander Shushunov, and John-Paul Teti for assistance in data preparation. We also thank Jordan Boyd-Graber for feedback and the University of

Maryland ProbMod reading group and CNL Lab for their helpful discussions. Parts of this work were previously presented at CMCL 2017 (Perkins, Feldman, & Lidz, 2017) and BUCLD 2017.

8 References

Abend, O., Kwiatkowski, T., Smith, N. J., Goldwater, S., & Steedman, M. (2017). Bootstrapping language acquisition. *Cognition*, *164*, 116–143.

Adriaans, F., & Swingley, D. (2012). Distributional learning of vowel categories is supported by prosody in infant-directed speech. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 34).

Alishahi, A., & Stevenson, S. (2008). A computational model of early argument structure acquisition. *Cognitive science*, *32*(5), 789–834.

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

Baker, C. L. (1979). Syntactic Theory and the Projection Problem. *Linguistic Inquiry*, *10*(4), 533–581.

Barak, L., Fazly, A., & Stevenson, S. (2014). Learning verb classes in an incremental model. In *Proceedings of the Fifth Workshop on Cognitive Modeling and Computational Linguistics* (pp. 37–45).

Brown, R. (1973). *A First Language: The Early Stages*. Cambridge, MA: Harvard University Press.

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

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.

Cameron-Faulkner, T., Lieven, E. V., & Tomasello, M. (2003). A construction based

analysis of child directed speech. *Cognitive Science*, *27*(6), 843-873.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American statistical association*, *83*(403), 596–610.

Fillmore, C. J. (1968). The case for case. In E. Bach & R. Harms (Eds.), *Universals in Linguistic Theory.* New York, NY: 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). Waltham, MA: Ginn and Company.

Fillmore, C. J., Kay, P., & O'connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language*, 501–538.

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., Gertner, Y., Scott, R. M., & Yuan, S. (2010). Syntactic bootstrapping. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(2), 143–149.

Fisher, C., Jin, K.-S., & Scott, R. M. (2019). The developmental origins of syntactic bootstrapping. *Topics in Cognitive Science*, 1–30.

Fodor, J. D. (1998). Parsing to learn. *Journal of Psycholinguistic Research*, *27*(3), 339–374.

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.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*(6), 721–741.

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.

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

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago, IL: University of Chicago Press.

Gomez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in cognitive sciences*, *4*(5), 178–186.

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). Cambridge, MA: MIT Press.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97–109.

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). Cambridge, MA: The MIT Press.

Hopper, P. J., & Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language*, 251–299.

Hudson Kam, C. L. H., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive psychology*, *59*(1), 30–66.

Jin, K.-S., & Fisher, C. (2014). Early evidence for syntactic bootstrapping: 15-month-olds use sentence structure in verb learning. In *Proceedings of the 38th Boston University Conference on Language Development.* Boston, MA: Cascadilla Press.

Keenan, E. L. (1976). Towards a Universal Definition of 'Subject'. In C. Li (Ed.), *Syntax and Semantics: Subject and Topic.* New York: Academic Press.

Kwiatkowski, T., Goldwater, S., Zettlemoyer, L., & Steedman, M. (2012). A probabilistic model of syntactic and semantic acquisition from child-directed utterances and their

meanings. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 234–244).

Landau, B., & Gleitman, L. R. (1985). *Language and Experience: Evidence from the Blind Child*. Cambridge, MA: Harvard University Press.

Langacker, R. W. (1999). *Grammar and Conceptualization*. New York, NY: Walter de Gruyter.

Lasnik, H. (1989). On certain substitutes for negative data. In R. J. Matthews & W. Demopoulos (Eds.), *Learnability and linguistic theory* (pp. 89–105). Dordrecht: Kluwer.

Lee, J. N., & Naigles, L. R. (2005). The input to verb learning in Mandarin Chinese: a role for syntactic bootstrapping. *Developmental Psychology*, *41*(3), 529.

Lee, J. N., & Naigles, L. R. (2008). Mandarin learners use syntactic bootstrapping in verb acquisition. *Cognition*, *106*(2), 1028–1037.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.

Levin, B., & Hovav, M. R. (2005). *Argument realization*. Cambridge: Cambridge University Press.

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

Lidz, J., & Gleitman, L. R. (2004b). Yes, we still need universal grammar. *Cognition*, *94*(1), 85–93.

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

MacWhinney, B. (2000). *The CHILDES project: The database* (Vol. 2). Psychology Press.

Maurits, L., Perfors, A., & Navarro, D. (2009). Joint acquisition of word order and word reference. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*.

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.

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

Naigles, L. R. (1996). The use of multiple frames in verb learning via syntactic bootstrapping. *Cognition*, *58*(2), 221–251.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive science*, *14*(1), 11–28.

Newport, E. L., Gleitman, H., & Gleitman, L. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: language input and acquisition.* Cambridge: Cambridge University Press.

Parisien, C., & Stevenson, S. (2010). Learning verb alternations in a usage-based Bayesian model. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 32).

Pearl, L., & Lidz, J. (2009). When Domain-General Learning Fails and When It Succeeds: Identifying the Contribution of Domain Specificity. *Language Learning and Development*, *5*(4), 235–265.

Pearl, L., & Sprouse, J. (2013). Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. *Language Acquisition*, *20*(1), 23–68.

Perfors, A., Tenenbaum, J. B., & Regier, T. (2006). Poverty of the stimulus? A rational approach. In *Proceedings of the Cognitive Science Society* (Vol. 28).

Perfors, A., Tenenbaum, J. B., & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of child language*, *37*(3), 607–642.

Perkins, L. (2019). *How grammars grow: Argument structure and the acquisition of non-basic*

syntax (Unpublished doctoral dissertation). University of Maryland College Park.

Perkins, L., Feldman, N. H., & Lidz, J. (2017). Learning an input filter for argument structure acquisition. In *Proceedings of the 7th Workshop on Cognitive Modeling and Computational Linguistics.*

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

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.

Resnik, P. (1996). Selectional constraints: An information-theoretic model and its computational realization. *Cognition*, *61*(1-2), 127–159.

Rowland, C. F., Pine, J. M., Lieven, E. V., & Theakston, A. L. (2003). Determinants of acquisition order in wh-questions: Re-evaluating the role of caregiver speech. *Journal of Child Language*, *30*(3), 609–635.

Schneider, J., Perkins, L., & Feldman, N. H. (2020). A noisy channel model for systematizing unpredictable input variation. In *Proceedings of the 44th Annual Boston University Conference on Language Development* (p. 533-547).

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.

Seidl, A., Hollich, G., & Jusczyk, P. W. (2003). Early Understanding of Subject and Object Wh-Questions. *Infancy*, *4*(3), 423–436.

Singleton, J. L., & Newport, E. L. (2004). When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive Psychology*, *49*(4), 370–407.

Soderstrom, M., Blossom, M., Foygel, R., & Morgan, J. L. (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language*,

35(4), 869–902.

Stromswold, K. (1995). The acquisition of subject and object wh-questions. *Language Acquisition*, *4*(1-2), 5–48.

Suppes, P. (1974). The semantics of children's language. *American Psychologist*, *29*(2), 103.

Suzuki, T., & Kobayashi, T. (2017). Syntactic Cues for Inferences about Causality in Language Acquisition: Evidence from an Argument-Drop Language. *Language Learning and Development*, *13*(1), 24–37.

Valian, V. (1990). Logical and psychological constraints on the acquisition of syntax. In L. Frazier & J. G. De Villiers (Eds.), *Language Processing and Language Acquisition.* Dordrecht: Kluwer.

Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition*, *40*(1-2), 21–81.

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

Yang, C. (2002). *Knowledge and learning in natural language*. Oxford: Oxford University Press.

Yuan, S., & Fisher, C. (2009, May). "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

Details of Gibbs Sampling

We use Gibbs sampling (Geman & Geman, 1984) to jointly infer T, ϵ, and δ, integrating over θ and summing over e, with Metropolis-Hastings (Hastings, 1970) proposals for ϵ and δ.

We begin by randomly initializing ϵ and δ, and sampling values of T for each verb given values for those input filter parameters. From observations of a verb with and without direct objects, the model determines which value of T was most likely to have generated those observations. For $k^{(v)}$ direct objects in $n^{(v)}$ sentences containing verb v, we use Bayes' Rule to compute the posterior probability of each value for $T^{(v)}$,

$$P(T^{(v)}|k^{(v)},\epsilon,\delta) = \frac{P(k^{(v)}|T^{(v)},\epsilon,\delta)P(T^{(v)})}{\sum_{T'^{(v)}} P(k^{(v)}|T'^{(v)},\epsilon,\delta)P(T'^{(v)})} \tag{7}$$

Bayes' Rule tells us that the posterior probability of a particular value of T given $k^{(v)}$ and the other model parameters is proportional to the likelihood, the probability of $k^{(v)}$ given that value of T and those parameters, and the prior, the probability of T before seeing any data. We assume that T is independent of ϵ and δ. In Simulation 1, we set a uniform prior over T, which is adjusted to reflect different biases about the proportions of transitivity categories in Simulation 2.

To calculate the likelihood, we must sum over e. This sum is intractable, but because all of the values of e for the same verb and the same direct object status are exchangeable, we make the computation more tractable by simply considering how *many* errors were generated for sentences with and without direct objects for a particular verb. We divide the $k^{(v)}$ observed direct objects for a verb into $k_1^{(v)}$ direct objects that were observed accurately and $k_0^{(v)}$ direct objects that were observed in error. The total $n^{(v)}$ observations for verb v are likewise divided into $n_1^{(v)}$ accurate observations and $n_0^{(v)}$ errorful observations. We then calculate the likelihood by marginalizing over $n_1^{(v)}$ and $k_1^{(v)}$, again assuming independence among T, ϵ, and δ,

$$p(k^{(v)}|T^{(v)},\epsilon,\delta) = \sum_{n_1^{(v)}=0}^{n^{(v)}} \left[\sum_{k_1^{(v)}=0}^{k^{(v)}} p(k^{(v)}|k_1^{(v)},n_1^{(v)},\delta) p(k_1^{(v)}|n_1^{(v)},T^{(v)}) \right] p(n_1^{(v)}|\epsilon) \tag{8}$$

The first term in the inner sum is equivalent to $p(k_0^{(v)}|n_0^{(v)},\delta)$, assuming we know $n^{(v)}$, the total number of observations for a particular verb. This is the probability of observing $k_0^{(v)}$ errorful direct objects out of $n_0^{(v)}$ errorful observations, which follows a binomial distribution with parameter δ,

$$p(k^{(v)}|k_1^{(v)},n_1^{(v)},\delta) = p(k_0^{(v)}|n_0^{(v)},\delta) = \begin{cases} \binom{n_0^{(v)}}{k_0^{(v)}} \delta^{k_0^{(v)}} (1-\delta)^{n_0^{(v)}-k_0^{(v)}} & \text{if } k_0^{(v)} \leq n_0^{(v)} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

The second term in the inner sum in (8) is the probability of observing $k_1^{(v)}$ accurate direct objects out of $n_1^{(v)}$ accurate observations, which follows a binomial distribution with parameter $\theta^{(v)}$,

$$p(k_1^{(v)}|n_1^{(v)},T^{(v)}) = \begin{cases} \binom{n_1^{(v)}}{k_1^{(v)}} (\theta^{(v)})^{k_1^{(v)}} (1-\theta^{(v)})^{n_1^{(v)}-k_1^{(v)}} & \text{if } k_1^{(v)} \leq n_1^{(v)} \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

Recall that $\theta^{(v)} = 1$ for the transitive category of T, and $\theta^{(v)} = 0$ for the intransitive category of T. For the alternating verb category, $\theta^{(v)}$ is unknown, so we integrate over all possible values of $\theta^{(v)}$ to obtain $\frac{1}{n_1^{(v)}+1}$.

The last term in (8) is the probability of observing $n_1^{(v)}$ accurate observations out of the total $n^{(v)}$ observations for verb v, which follows a binomial distribution with parameter $1-\epsilon$,

$$p(n_1^{(v)}|\epsilon) = \binom{n^{(v)}}{n_1^{(v)}} (1-\epsilon)^{n_1^{(v)}} (\epsilon)^{n^{(v)}-n_1^{(v)}} \tag{11}$$

After sampling values for T for each verb in the dataset, we then sample values for ϵ

and δ. If T denotes the set of values $T^{(1)}, T^{(2)}, ..., T^{(V)}$, and k denotes the full set of observations of direct objects $k^{(1)}, k^{(2)}, ..., k^{(V)}$ for all V verbs in the input, we can define functions proportional to the posterior distributions on ϵ and δ, $f(\epsilon) \propto p(\epsilon|T, k, \delta)$ and $g(\delta) \propto p(\delta|T, k, \epsilon)$, as

$$f(\epsilon) = p(k|T, \epsilon, \delta)p(\epsilon) \tag{12}$$

$$g(\delta) = p(k|T, \epsilon, \delta)p(\delta) \tag{13}$$

where the likelihood $p(k|T, \epsilon, \delta)$ is the product over all verbs v of $p(k^{(v)}|T^{(v)}, \epsilon, \delta)$, as calculated in (8).

Within the Gibbs sampler, we resample ϵ using 10 iterations of a Metropolis-Hastings algorithm. We begin by randomly initializing ϵ. At each iteration, we propose a new value ϵ', sampled from the proposal distribution $Q(\epsilon'|\epsilon) = N(\epsilon, 0.25)$. Because the proposal distribution is symmetric, this new value is accepted with probability

$$A = min\left(\frac{f(\epsilon')}{f(\epsilon)}, 1\right) \tag{14}$$

If the new value ϵ' has higher probability given T, k and δ under equation (12), it is accepted. If it has lower probability under equation (12), it is accepted at a rate corresponding to the ratio of its probability and the probability of the old value of ϵ. After sampling ϵ, we resample δ with 10 iterations of Metropolis-Hastings. The proposal and acceptance functions are analogous to those for ϵ.

We ran multiple chains from different starting points to test convergence of T, ϵ, and δ. For the simulations reported here, we ran 1,000 iterations of Gibbs sampling. We took every tenth value from the last 500 iterations as samples from the posterior distribution over T, ϵ, and δ.