Original Article

Language *and* Speech

Language *and* Speech
1–33

DOI: 10.1177/00238309211066086
journals.sagepub.com/home/las
SAGE

Piecewise Structural Equation Modeling of the Quantity Implicature in Child Language

John Grinstead
Pedro Ortiz-Ramírez
Ximena Carreto-Guadarrama
The Ohio State University, USA

Ana Arrieta-Zamudio
The University of British Columbia, Canada

Amy Pratt
University of California, Irvine, USA

Myriam Cantú-Sánchez
Universitat de Barcelona, España

Jonathan Lefcheck
The Tennenbaum Marine Observatories Network, MarineGEO, Smithsonian Environmental Research Center, USA

David Melamed
The Ohio State University, USA

Abstract

We review an array of experimental methodological factors that either contribute to or detract from the measurement of pragmatic implicatures in child language. We carry out a truth value judgment task to measure children's interpretations of the Spanish existential quantifier *algunos* in implicature-consistent and implicature-inconsistent contexts. Independently, we take measures of children's inhibition, working memory, attention, approximate number ability, phrasal syntax, and lexicon. We model the interplay of these variables using a piecewise structural equation model (SEM), common in the life sciences, but not in the social and behavioral sciences. By 6 years of age, the children in our sample were not statistically different from adults in their interpretations. Syntax, lexicon, and inhibition significantly predict implicature generation, each

Corresponding author:
John Grinstead, The Ohio State University, 298 Hagerty Hall, 1775 College Road, Columbus, OH 43214, USA.
Email: grinstead.11@osu.edu

accounting for unique variance. The approximate number system and inhibition significantly predict lexical development. The statistical power of the piecewise SEM components, with a sample of 64 children, is high, in comparison to a traditional, globally estimated SEM of the same data.

Keywords
Implicatures, algunos, Spanish, piecewise structural equation model

Introduction

What domains of cognition must interact during cognitive development to allow children to grasp that the sentence in 1 is not pragmatically felicitous?

1. #Algunos niños bajaron la resbaladilla. (4 of 4 children in the scene
   Some children went down the slide. have gone down a slide)

Neo-Gricean pragmatics holds that this utterance is truth-conditionally correct, in that, if *all* the children under consideration went down a slide, then it is also true that *some* children did. However, the inference, or scalar implicature produced by the sentence in 1, to the effect that "some, but not all" children went down the slide, is nonetheless incompatible with the context, rendering the sentence pragmatically infelicitous.

There seems little doubt that to understand this sentence, development in the lexicon is critical. The existential quantifier *algunos* comes to have its "some, but not all" meaning by virtue of its relationship to *todos* (all). The quantifier in subject position, however, does not occur in isolation. Rather, it occurs as a syntactically concatenated lexical item together with a noun to form the *algunos niños* (some children) noun phrase constituent. This noun phrase has been further syntactically merged, as the subject of the entire sentence, with a predicate, to produce the sentence. Thus, syntax must also play a role. The "some, but not all" implicature interpretation of *algunos* is not its only possible interpretation, however, as it could also be interpreted to mean "some, and possibly all." Thus, the domains of cognition typically taken to play a role in sentence disambiguation by natural language processing research, namely, executive function (EF), must also play a role. Finally, an existential quantifier, as the name suggests, conveys quantity information. It seems highly unlikely that human languages express this quantity information in the lexicon without the number domain of cognition also playing a role.

An active literature over the past 20 years has debated the questions of how much children know about these interpretations, when they know it, and how this knowledge varies in its cross-linguistic instantiations. Nonetheless, the developing system of conceptually necessary components of cognition that must interact to generate scalar implicature interpretations has not been modeled. We attribute this lacuna to a plethora of methodological obstacles, which we will review, on one hand. On the other hand, however, even when implicature generation is measured accurately, sample size can be prohibitive with respect to the allowable complexity of and ultimate inference provided by such models. To evaluate the system of cognitive variables that produces these interpretations, statistical techniques, such as globally estimated structural equation model (SEM) may be used. However, to power such a model, a very large-scale study is required, with 300 or so children, which is often infeasible. In what follows, we suggest that studies of such scale may not be mathematically necessary for statistically valid modeling of these systems of variables to occur.

In this study, we review what we discern to be the obstacles as well as the boons to the measurement of children's knowledge of scalar implicatures. Based on this review, we then attempt to implement what we take to be the best practices and to avoid the confounds we aware of for scalar implicature measurement with a cross-sectional sample of 4- to 8-year-old children and adult controls. In addition to our implicature outcome variable, we independently measure lexicon, syntax, EF, and non-linguistic number. We then model these variables in the traditional way, using multiple regression in a generalized linear model. In what we hope could be a contribution to more tractable measurement of systems of behavioral scientific variables, we then compare this flat-structured multiple regression to a hierarchically structured piecewise SEM, following the work of William Shipley (2000, 2002, 2009) in the natural sciences. We believe that this approach to modeling systems, which emerges from the modeling of ecosystems in the life sciences, may be of use to those attempting to model cognitive systems, as well.

2 Methodological heterogeneity

In the literature on the development of children's knowledge of the quantity implicature, there is a great deal of methodological heterogeneity, making it hard to draw firm conclusions about what children know. In what follows, we identify sources of heterogeneity and attempt to discern methodological choices that are most likely to provide insight into children's emerging knowledge of pragmatic implicatures.

2.1 Generics

A prominent obstacle to measuring what children know about scalar implicatures has been the use of generic statements, formulated using individual-level predicates (Carlson, 1977), in experimental sentences. Such generic statements attribute a trait, characteristic or property to an individual or a situation and children are asked whether these traits or properties hold of "some" versus "all" of a particular class of entities. Examples of such generics are the following stimuli items from Noveck (2001, p. 187—his translation from French, based on Smith, 1980).

1. Some giraffes have long necks.
2. Some televisions have screens.
3. Some cars have motors.
4. Some cats have ears.
5. Some airplanes have wings.

The logic of the experiment suggests that these prototypical properties should lead a participant to regard them as holding of all such objects. Thus, when participants are asked if they hold "some" of these objects, the expected implicature-driven answer is "no," with the rationale that "all" such objects have that trait. Alas, the experiment is only useful for measuring implicature-generating ability if we are confident that children and adults have categorically unambiguous judgments of the property being associated with the entity.

Eight- to 10-year-old child French speakers struggled to give "no" answers to the question of whether the sentences in 1 to 5 were true. Noveck (2001) reports that they only rejected the sentences 11% of the time. Worse still, adults rejected them only 59% of the time. The difference between the adult answers and the child answers was significant; however, this significant difference might have had more to do with child versus adult knowledge of the world and of what traits are plausible for particular objects and creatures, than it did with knowledge of implicatures.[1]

In work using similar generic sentences, but in English, Feeney et al. (2004) were able to get children to reject existentials at virtually the same 66% rate as adults, making it seem unlikely that the difference in earlier work was attributable to pragmatic development, though there could obviously have been language-specific differences, too. The fact that adult judgments were not categorical here either, however, raises the question of what adults thought the task was about. In this way, the use of a sentence verification task with generic statements appeared to prevent adult and child participants from demonstrating what they knew about pragmatic implicatures. This is particularly the case, given the less-than-categorical adult responses.

2.2 Overt versus covert partitives

Using individual-level predicate generics has contributed one obstacle to understanding what children know about pragmatic implicature interpretations and the indiscriminate use of both overt and covert partitives in this literature has contributed a second confound. These two construction types have different semantic representations and consequently interact with linguistic pragmatics in different ways. While each construction is interesting for different reasons, treating them as if they were the same muddles our understanding. At the level of semantics, the difference between the phrases in 6 and 7 relates to domain restriction, as discussed by von Fintel (1994) and Roberts (1995), whom we follow:

6. Overt partitive—some of the dwarves

7. Covert partitive—some dwarves

The presence of the definite article *the* in 6 invokes Russell's (1905) uniqueness presupposition which conveys to the listener that the speaker presupposes that there is a unique set of dwarves that is specific to both the speaker and the hearer. The presupposition allows no ambiguity and the indefinite exemplars referred to by *some* must be drawn from this restricted domain. In contrast, in the covert partitive in 7, there is no presupposition restricting the domain from which the dwarves are drawn. Consequently, the phrase could refer to a subset of all dwarves that fall under the scope of the existential *some* (all the dwarves that exist), or, if the listener has developed adult pragmatic competence, the listener may restrict the domain of dwarves to those currently under discussion in a given context, such as a typical adult conversation, or in our case, in the experimental context story presented in a truth value judgment task (TVJT). Thus, domain restriction is the critical difference between these two partitive types. To interpret a covert partitive, one must make a pragmatic decision about which of two possible domains is relevant: just any plural group of dwarves out of the universe of possible dwarves or a plural group of dwarves that is a subset of dwarves that are prominent in the conversational common ground (Stalnaker, 1974), as in a pragmatic implicature interpretation. In contrast, to interpret the overt partitive in 6, the domain is restricted by presupposition to a set of dwarves that the speaker takes the listener to be familiar with. There is no decision to be made by the listener as to the specificity of the domain from which the indefinite exemplars modified by *some* is to be drawn.[2]

From this perspective, any experiment using overt partitives is testing children's knowledge of domain restriction via the uniqueness presupposition, *in addition* to testing what they know about indefinite quantifier interpretations. Experiments using covert partitives, in contrast, test what children know about distinguishing domains that are pragmatically restricted by the question under discussion (QUD) to the set of dwarves that are prominent in the common ground from the domain of all possible dwarves, and only then, testing what children know about indefinite quantifier interpretation. In adult semantics, at least, there is no decision to make about the domain from which the indefinite is drawn in overt partitives, assuming knowledge of the uniqueness presupposition, while there is a decision to be made in covert partitives about which domain is relevant, given that adults frequently do use a covert partitive to refer to any plural number of members of a set.

In sum, it is helpful in science to make theoretical claims using the same terms to refer to the same constructs. If we want to theorize about scalar implicatures, using quantifiers drawn from the quantity scale, it is helpful to refer to the same syntactic and semantic construction, which it should be clear that we have not. In what follows, we will follow researchers in the field who have made claims about covert partitives, including Chierchia et al. (1998), Guasti et al. (2005), Miller et al. (2005), Pouscoulous et al. (2017), and Vargas-Tokuda et al. (2009), as in 8–11.

8. Some dwarves went for a ride in the boat. (Chierchia et al., 1998, p. 104)

9. Some girls are watching TV. (Guasti et al., 2005, p. 696)

10. Je voudrais que quelques boîtes contiennent un jeton.
    I would like some boxes to contain a token. (Pouscoulous et al., 2017, p. 361)

11. Algunos tigres saltaron sobre King Kong.
    Some tigers jumped over King Kong. (Vargas-Tokuda et al., 2009, p. 110)

The results of research studying children's interpretations of indefinites in overt partitive constructions (e.g., Huang & Snedeker, 2018; Katsos & Bishop, 2011; Papafragou & Musolino, 2003) may be interesting, but ultimately are not telling us anything about children's ability to choose the domain from which those indefinites are drawn, following von Fintel (1994) and Roberts (1995), which we take to be a central aspect of developmental pragmatics.

2.3 English "some" and phonological effects

A further confound in the literature comes from the fact that *some* in English may vary by, at least, vowel duration and pitch in ways that correspond systematically to its interpretation. Three of the most prominent variants of *some* are the vowel-less version, [sm] (see Milsark, 1977; Postal, 1964), the full-vowel, non-pitch-accented version [səm], and the full-vowel, pitch-accented version [sʌm]. The pitch-accented version occurs with an L + H* contour, which is associated with nuclear stress, in the terms of autosegmental metrical phonology (Beckman & Pierrehumbert, 1986; Pierrehumbert, 1980). Thorward (2009) and Grinstead et al. (2010) show that for adults, [sm] tends to occur with the "some, and possibly all" meaning of the existential, pitch-accented [sʌm] tends to occur with the "some, but not all" pragmatic implicature version of the existential and that the full-vowel, non-pitch-accented [səm] can yield either interpretation. The problem this poses for studies of existential quantifier interpretation that purport to study what children know about these quantifiers as a function of semantic (e.g., downward-entailing) context is that simply knowing the phonetic forms of these quantifiers is likely very influential, if not sufficient by itself, independent of syntactic context, to know whether they are associated with the pure existential versus the pragmatic meaning. Because most of the work in the literature does not control this variable, it is unclear whether their experiments demonstrate children's knowledge of the interaction of syntactic context and pragmatic implicatures or whether they simply demonstrate that children have learned specific sound-meaning correspondences. These problems do not arise in Spanish, which does not have an interaction of vowel duration and prosody of the same character as English on quantifiers (see Astruc et al., 2012). To avoid this confound, we study the development of the quantity implicature in child Spanish.

3 Factors that facilitate measurement of the quantity implicature

As we have tried to make clear, generic sentences are not the most appropriate type for testing children's knowledge of the quantity implicature. Sentences with eventive predicates, in contrast, make it possible for children to judge whether some, or all, participants have carried out an activity that has just been observed in a TVJT scenario. Furthermore, it may be independently interesting to study what children know about overt partitive structures that include definiteness presuppositions in them, for example, the elephants pushed some of the trucks. However, the overt partitive structure is different from the covert partitive and introduces a set of additional variables, which could make it harder to measure just what children know about the quantity implicature, which is our goal. Finally, the integration of knowledge of how pitch and duration interact with the existential quantifier *some* in English is interesting, but again, represents an unnecessary complication to understanding what children know about the implicature. While Spanish *algunos* can be prosodically focused to derive an implicature interpretation, it need not be. In the absence of prosodic focus, the implicature persists, ceteris paribus, making child Spanish a good choice for studying the phenomenon of interest.

3.1 "algunos"

The plural existential Spanish determiner *algunos* (Gutierrez-Rexach, 2001; López-Palma, 2007; Martí, 2008) allows contrastive focus to be expressed prosodically, similar to English. However, this prosodic function is not a necessary component, as it appears to be in English, for the generation and cancelation of implicatures associated with different variants of *some*. Rather, *algunos* encodes specific semantic properties lexically, independent of prosody, which makes studying logical versus pragmatic uses of the existential more straightforward. Concretely, in previous work, it was possible to study what preschool children know about existentials in Spanish (e.g., Miller et al., 2005; Vargas-Tokuda et al., 2009), without considering the confounding effects of phonological development in English.

algunos can be linked to the conversational common ground, and consequently can generate a pragmatic implicature (Gutierrez-Rexach, 2001, 2010), which is illustrated in 12. Quantity implicatures are generally canceled in downward-entailing environments, such as the antecedent clause of a conditional sentence. Thus, in 13, *algunos* no longer has its "some, but not all" implicature and instead gets a logical "some, and possibly all" interpretation.

<u>4 Out of Our 4 Cats Are In a House</u>

12. #Algunos gatos están en la casa. ("some, but not all" infelicitous)
    Some cats are in the house

13. Si algunos gatos están en la casa, me das una moneda.
    If some cats are in the house, you give me a coin.

In Vargas-Tokuda et al. (2009), the 5-year-old monolingual participants were able to generate quantity implicatures with *algunos* 70% of the time and adults did so 80% of the time. While this seemed categorical, there was only one predicate, "Algunos X jumped over Y.," which limited the generality of the claim, and the tasks were live, which made for a less reliable presentation than would a video-recorded TVJT. In the current project, our design attempts to overcome both of these shortcomings, among others.

3.2 Structuring the discourse—an explicit QUD

In the work of Pratt et al. (2018), it is shown that using eventive predicates in Spanish with covert partitives and an explicit QUD in their TVJT, highly categorical judgments can be elicited from adults with respect to quantity implicatures. QUD is a construct discussed by Roberts (2003), Clifton and Frazier (2012), and others, which is thought of as an implicit question to which all conversational contributions are a response (e.g., Collingwood, 1940; Groenendijk & Stokhof, 1984). In this way, the QUD structures the flow of discourse and can be used to distinguish utterances as more or less relevant to the topic of conversation, making the relationship between a sentence and a pragmatic context clearer. Gualmini et al. (2008) showed that effective use of the QUD in TVJTs (e.g., their question–answer requirement), measuring children's quantifier scope interpretations, was sufficient to produce adult-like scope judgments of scopally ambiguous sentences, which is normally challenging for children. Musolino (1998) refers to this challenge or developmental obstacle for children as the observation of isomorphism.

Following Gualmini et al. (2008), Pratt et al. (2018) structured the discourse of their TVJTs measuring *algunos* interpretation to make it clear that the experiment was about "how many" participants or "who" among the participants would engage in a particular action. This manipulation did not result in children generating pragmatic implicatures at high levels in this experiment. In fact, only 8% of *algunos* trials in implicature-generating contexts were rejected (rejection indicates implicature use in this experiment) by 5-year-old monolingual Spanish-speaking children (n=42). However, adult Spanish speakers, rejected the same trials 94%–100% of the time, showing implicature generation rates that were much higher than adult implicature generation rates in the contemporary literature. On this basis, the authors concluded that use of an explicit QUD in structuring the discourse of a TVJT is a valuable experimental design feature, yet not sufficient to elicit implicature interpretations from children. Critically for our next section, in these scenarios, the action depicted in the TVJTs corresponding to the eventive predicates, such as 14, was simultaneous, not consecutive.

14. Algunos cerditos rescataron al caballo.
    Some piggies saved the horse.

That is, in the scenario associated with sentence 14, for example, the piggies (Peppa Pig and her friends) all moved together, simultaneously, to rescue the horse, which had gotten stuck among some rocks. This type of simultaneous action occurred in all of the other scenarios, as well. In later work, Pratt et al. (2019) hypothesized that this design choice had serious consequences for the children's ability to interpret the simultaneous action TVJT scenario, which was changed in the subsequent version of the experiment produced by these authors.

3.3 Working memory and implicatures—consecutive versus simultaneous action

In the work of Pratt et al. (2019), it was shown that by changing the action in the TVJT videos from simultaneous to consecutive, maintaining an explicit QUD, with eventive predicates, adult categorical (96%) implicature generation could be sustained. However, in this TVJT, the children, who ranged in age from 4- to 8-year old, also generated implicatures at 80%. In fact, there was a strong effect of age, and while adults were significantly different in acceptance from 4- and 5-year-olds, they were not different from 6-, 7-, or 8-year-olds. Why should there be such a difference between simultaneous versus consecutive action in the TVJTs for children's implicature generation?

Studies in the field of attention and memory have consistently shown effects of "retention interval duration" in adults (Morey & Bieler, 2013; Ricker & Cowan, 2010; Woodman et al., 2012; Zhang & Luck, 2009). The longer that simultaneously presented stimuli are held in working memory, the worse the retention. Studies that present items in a consecutive fashion do not show effects of retention interval duration (Barrouillet et al., 2004; Lewandowsky et al., 2004). That is, there is no relationship between the length of time between stimulus presentation and how well that information is retained. Ricker and Cowan (2014) show that the time interval between stimulus presentations in consecutive presentation experiments allows for memory consolidation. This is the key difference between most simultaneous versus consecutive presentation experiments. These authors manipulate the time available for memory consolidation across experiment types and make the difference in forgetting disappear.

Although there is no tightly controlled comparison in implicature generation experiments that only manipulates the time available for memory consolidation in children that we are aware of, we can say that with relatively few changes in design between the task in Pratt et al. (2018) and the task in Pratt et al. (2019), results went from 8% implicature generation in 5-year-olds to 74% implicature generation in an age-matched sub-group of the 2019 participants. Further confirmation that working memory is clearly relevant to the difference, following Ricker and Cowan (2014), comes from the fact that their dot counting auditory working memory measure (Case et al., 1982) significantly predicted implicature generation in the 2019 sample ($r^2 = .163$, $p < .001$).

4 A theory of language, cognition, and the quantity implicature

We have described an array of measures that are relevant to effectively measuring the quantity implicature. What relationship does the quantity implicature have to distinct domains of language and cognition?

4.1 Lexical refraction—number and scalar development

In the work of Grinstead et al. (2019), it was shown that implicature generation with *algunos* was predicted by both a receptive measure of lexicon, the Test de Vocabulario en Imágenes Peabody (TVIP; $r^2 = .473$, $p < .001$, $n = 54$), and by an expressive measure of lexicon, the *Adivinanzas* ("Riddles") subtest of the Batería de Evaluación de Lengua Española (BELE; $r^2 = .220$, $p < .001$, $n = 54$). This finding is consistent with the lexical refraction hypothesis of Grinstead et al. (2021).

Lexical refraction claims, first of all, that the approximate number system (ANS) provides quantity sensitivity to the natural language lexicon, which refracts this basic distinction as a function of each quantity-related morpheme's semantics. ANS is thought of as a non-verbal ability not unique to humans, but rather one that is shared across species. It allows the representation of numerosities and magnitudes, the accuracy of which depends on numerical ratios (e.g., Agrillo et al., 2012; Dehaene, 1997; Feigenson et al., 2004; Xu & Spelke, 2000). This fundamental, non-verbal representation of quantity would be refracted as cardinality for numeral quantifiers; plurality for nominal, adjectival, determiner, or verbal number marking morphemes in Spanish; trial number markers in languages, such as Egyptian Arabic, which mark "exactly three" as a nominal morpheme; distributive, as in *each*, and so on. Second, it claims that the relationships among natural language quantifiers on a lexical pragmatic scale (the quantity scale, in this case) must be learned one-by-one and the links among the members of the scale grow gradually and implicitly with greater exposure and contrastive use. On this hypothesis, the overall size of the lexicon should index the strength of children's ability to, in the case of the quantity implicature, negate the stronger members of the scale, which would yield the "some, but not all" interpretation. In the work of Pratt et al. (2020), it is further shown that four distinct lexical measures can identify *algunos* quantity implicature generators and quantity implicature non-generators with 91% sensitivity and 100% specificity in a sample of 52 Spanish-speaking children, using a linear discriminant function analysis. In short, the lexicon plays a major role, perhaps unsurprisingly, in the development of children's abilities to draw inferences from pragmatic scales of lexical items.

4.2 Syntax—compositional semantics

As we have seen, the lexicon appears to play a major role in the generation of scalar implicatures. However, children and adults are obviously drawing these interpretations not from isolated quantifiers, but rather from quantifiers that have been concatenated with other lexical items and syntactically composed constituents into larger syntactic structures. Although a great deal of children's syntactic knowledge seems adult-like by the time they are 4 years old, which is the age of the youngest children in our sample, there is, nonetheless, morphosyntactic and phrasal syntactic development that takes place in this preschool to school-aged sample. For example, our mean length of utterance in words (MLUw) measures, calculated from two distinct spontaneous production samples (one unstructured and one a Frog Story), both significantly correlated with age in months in our sample,[3] described in greater detail below.

We tend to think of syntax as being largely adult-like by the age of our participants, but an influential formal syntactician has argued that there is a class of exceptions to this generalization. Reinhart (1998, 2004) speculated that the interaction or interface that occurs between pragmatics and syntax in *algunos*-type sentences created variance, or apparently non-adult-like behavior, in child interpretations of scalar implicatures, stress shift, focus constructions, and others. We follow this prescient observation and respond to it by adding an independent measurement of syntactic ability to our protocol. If no statistically significant relationship between *algunos* sentence interpretation and independently measured syntactic development (our ARG-I, below) obtains, then the observed development must be external to syntax. If, however, the covariance relationship between our syntax measure and our sentence interpretation measure is statistically significantly related, then we must conclude that development occurs in the domain of syntax itself and that it contributes to the interpretation of sentences that can express pragmatic implicatures. It does not seem difficult to imagine that a linguistic ability could be present, but not yet mature enough, akin to adult-like levels, to allow a robust interface with linguistic pragmatics, as in the case of scalar implicature interpretations, following Reinhart's argument.

In fact, we use interpretation of relatively unambiguous and pragmatically less-challenging sentences as filler sentences in our protocol—effectively ensuring a minimum level of syntactic competence to even be included in the protocol. As usual, filler items also ensure that children are able to pay attention to the task and that they are oriented to what the task is about. As we detail below, children must correctly interpret filler sentences with *todos* ("all") and *ningún* ("none") in subject position to be included in our sample. If variance in syntactic ability is detectably contributing to quantifier interpretation, then this variance should be visible in sentences with *algunos* in subject position.

It is worth noting that syntax must also play a critical role in the adult implicature interpretation of existential quantifiers. However, because adults presumably show much less variance in their syntactic abilities than do children, it is much harder to statistically model the role the syntax plays in these interpretations. This is yet another advantage of working with child language, inasmuch as incomplete development of the components that theoretically determine interpretation allow for variance to aid in our understanding of the interaction of these components. The role of syntax in implicature interpretations, to our knowledge, has not been tested, as of yet, in the existing literature.

4.3 EF—ambiguity resolution

There is substantial evidence that domain-general inhibition plays a role in matching concepts to candidate lexical items.[4] In particular, there is a significant developmental relationship between the lexicon and inhibition. A way of thinking about this is that the more lexical items there are in semantic proximity to a concept, the more competitors there will be that need to be inhibited, to arrive at the target (see Gangopadhyay et al., 2019; McMurray et al., 2019 inter alia). Furthermore, another domain of EF, auditory working memory (see Houston et al., 2020 for review), is also strongly predictive of lexical development. Here, the conceptual connection is between being able to move phonological information encoding lexical items into long-term memory, as a function of one's capacity to retain this information in working memory.

It is also true, however, that the full sentences in which lexical items, including existential quantifiers, occur can be ambiguous. Such sentences should consequently require EF abilities to be processed in an adult-like way, as assumed by natural language processing research (see Novick et al., 2005 for review). That is, once lexical items are chosen to express a concept or function and are then composed by syntax into a sentence, EF must determine whether such a sentence is appropriate for a pragmatic context or not. Here, we see inhibition, at least, playing an additional role in allowing or disallowing the interpretation of a sentence in a given situation.

We assume the model of EF of Miyake et al. (2000) and further that there is an independent developmental trajectory for its components, which include auditory working memory, attention, and inhibition. Each of these variables has been shown to be predictive of quantity implicature interpretations, cross-linguistically in adults and children (e.g., Antoniou et al., 2016; De Neys & Schaeken, 2007; Janssens et al., 2014; Kapa & Colombo, 2014; Marty et al., 2013; Wang, 2019) and has been shown to be predictive specifically with *algunos* in the work of Grinstead et al. (2019).

5 SEM and cognition

Given the predictor variables we have considered, a standard means of modeling their relationships to our implicature generation outcome variable would be to use a multiple regression, which can be represented graphically using a box-and-arrow diagram as in Figure 1.

While a multiple regression model would be the conventional way to proceed, the theory we propose for the processing of implicature interpretations is not unstructured, as the diagram in Figure 1 implies. Rather, following the lexical refraction hypothesis, we expect that the ANS delivers quantity information to the lexicon, and not directly to fully formed sentences. On this view, the sentences that carry existential quantifiers provided by the lexicon must have also been syntactically composed from these and other lexical items that were appropriate to the specific concepts and intentions that the sentence is created to convey. In this way, we expect that syntax, lexicon, and EF work together to generate a quantity implicature. Moreover, domain-general EF seems likely to play a role not only in determining the pragmatic appropriateness of sentences for contexts, following natural language processing research, but also seems likely to be aiding in the

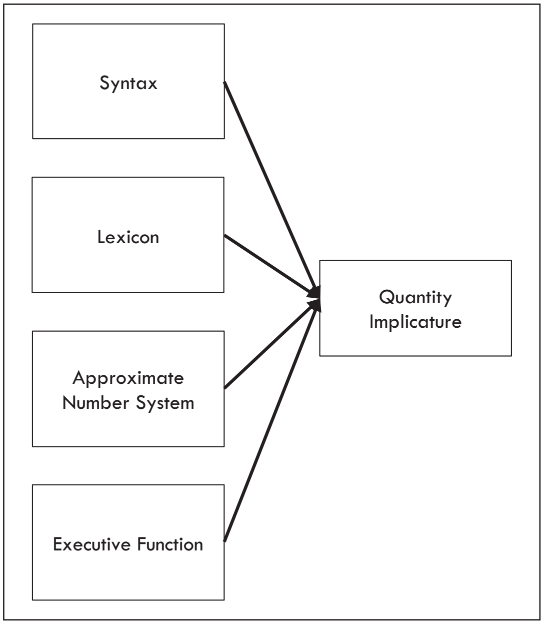

Figure 1. Multiple regression model with syntax, lexicon, the ANS, and EF as predictor variables, and quantity implicature interpretations as the dependent variable.

choice of lexical items when they interface with concepts, following much research in developmental psychology and speech and hearing science (e.g., McMurray et al., 2019). We expect EF to be playing both a direct and an indirect role (through lexicon) in the processing of quantity implicatures, perhaps as in Figure 2.

A model that allows for the consideration of both direct and indirect effects that allow observed variables to function as both predictors *and* responses is possible through the technique of SEM. In the traditional application of SEM in the social and behavioral sciences, covariances among all variables (both predictors and responses) are estimated simultaneously. Because of this, increasingly complex models require a very large number of participants to adequately reproduce the global variance–covariance matrix associated with them (Kline, 1998; Wolf et al., 2013). However, over the last 20 years, there has been an increasing movement to integrate more graph-theoretic approaches into SEM, motivated primarily by a desire to speak to causality (i.e., the directional arrows in Figure 2, which are interpreted as one variable leading to, or causing, another, e.g., Pearl, 2009). The merger of graphical theory and SEM has loosened a number of restrictions that plagued earlier generations of SEM, including sample size, but also permitting more complex variance structures (i.e., random effects) and non-parametric (i.e., non-Gaussian) responses. This type of SEM, referred to as a piecewise SEM, has found considerable traction in the life sciences, and in the field of ecology, in particular. The reasons for this adoption are intimately related to the circumstances in which ecological data are collected. To appropriately model a particular ecosystem, there may simply not be hundreds of exemplars of the particularly configured elements under investigation to model (e.g., light, temperature, nutrients, microbes, invertebrates). Nonetheless, the field, in particular, the work of Shipley, Grace, Schoolmaster, and Lefcheck, has shown that it is possible to model the hypothetically causally structured relationships among a set of variables.

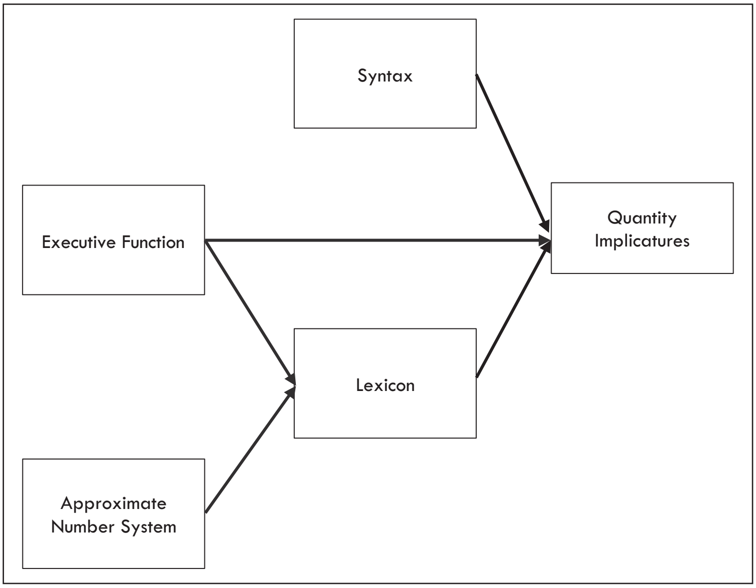

Figure 2. SEM with syntax, lexicon, and EF as direct predictor variables of the quantity implicature, with EF and the ANS predicting the lexicon.

How does piecewise SEM work, exactly? In contrast to globally computed SEM, which estimates all relationships simultaneously, in locally estimated SEMs (e.g., Grace et al., 2015; Lefcheck, 2016, 2019; Schoolmaster et al., 2020; Shipley, 2000, 2002, 2009), each response (or endogenous) variable is modeled independently. The primary benefit of this is that power for a given path is based on the information *in that path*, rather than based on the variance/covariance matrix for the entire system of paths. That is, statistical power in such a model is determined for each local relationship's regression model, which implies that a model with three local regression relationships will have three separate variance–covariance matrices, which obviously require less power than a single matrix with all of the relationships represented in it simultaneously. While this approach has some drawbacks (see below), it maintains key generalizations of SEMs, including modeling non-continuous endogenous variables, hierarchical or nested non-independent observations, and different estimation procedures (e.g., maximum likelihood vs. least-squares), all of which better address the underlying processes generating the data commonly reported in both ecological and social sciences.

In piecewise SEM, the directed relationships among variables are determined a priori by the investigator to reflect a single multivariate causal hypothesis (see Figure 2). The justification for these paths as causal can arise from many sources, including logical constraints, prior evidence or experimentation, and theory. The investigator can then challenge the supposed causal relationships with actual data to support or refute the original hypothesis. To determine whether the model fits the data, we conduct tests of "directed separation," which statistically evaluate missing linkages that were not a priori included on the basis of being deemed causally unimportant. If one or more of these linkages turn out to be significantly statistically correlated, then it would imply that the hypothesized structure is missing key linkages inferred by the data, and therefore must be rejected. Each missing linkage is referred to as an independence claim (with the null hypothesis being that the two variables under consideration are indeed independent, conditional on the relationships already specified in the model), and the collective independence claims form what is known as the basis set.

The tests of directed separation can be summarized into a single Fisher's C statistic, which is computed using the following formula of Shipley, given in the work of Lefcheck (2016).

$$C = 2\sum_{i=1}^{k} ln(p_i)$$

In this formula, k is the number of independence claims in the basis set, i is the ith claim (claim of independence among variables in a possible path), and p is the p-value from the corresponding significance test. Lefcheck explains that, ". . . Shipley showed that C is χ^2-distributed with 2k degrees of freedom. Thus, a model-wide p-value can be obtained by comparing the value to a χ^2 table with the appropriate degrees of freedom." In this way, C is analogous to the χ^2 statistic used to evaluate fit in earlier generations of SEM.[5]

In Shipley's (2009, p. 365) terms, "Compare the resulting C value to a chi-squared distribution with 2k degrees of freedom. Reject the causal model if the C value is unlikely to have occurred by chance (i.e. below the chosen significance level)." This is because the null hypothesis of the C-test is that the model fits the data. A significant result implies that the path-analytic model we specified does not fit the data. In terms of this specific test, the rejection of the model suggests two or more variables are not conditionally independent, that is, they are significantly correlated beyond what might be expected by chance. Conversely, an *insignificant* result means that we cannot reject the hypothesis that our model fits the data. This implies that a p-value of greater than, for example, .05 for the behavioral sciences, is an indicator of good fit. We mention this because aspiring to p-values of above .05 is somewhat counter-intuitive in the behavioral sciences.

Thus, the p value of less than .05 in this case means that an alternative causal hypothesis is probable, and we can use the significance values in the tests of directed separation to identify those paths, which may—depending on the goal—be introduced into the model to improve fit, as long as they are causally justified. To reiterate, a p-value of greater than .05 is desirable because it implies that alternate model configurations—for example, those including paths that were formerly lacking in the proposed model, but which were identified as significant in the tests of directed separation—are probable within a given threshold for type I error (typically $\alpha = .05$).

Thus, the basis set for the model in Figure 2 includes test of the significance of a path in which lexicon is predicted by syntax + EF + the ANS and another in which quantity implicatures are predicted by syntax + EF + the ANS + lexicon. All of these alternative relationships are tested and if any of them are significant, it reduces the probability that the proposed model structure reproduces the relationships implied by the data. Lefcheck (2016, 2019) has written an R package ("PiecewiseSEM"), which carries out these computations and calculates both the Fisher's C statistic, with appropriate degrees of freedom, and reports whether any of the alternative paths in the basis set are significant.

To be clear, piecewise SEM is not a panacea. For example, it does not yet incorporate latent or composite variables, which is one of the great advantages of conventional SEM.[6] It must also meet all the assumptions of the underlying (generalized) linear models, such as homogeneity of variance. However, for behavioral science in general, and linguistics, in particular, it holds out the prospect of being able to model a network of multiple causally dependent variables in a systemic fashion without incurring the exorbitant expense of running a multi-task protocol with over 200 or 300 participants.

6 Research questions

Summarizing, there are a number of confounds to accurately ascertaining what children know about quantity implicatures. In spite of these challenges, we have identified other factors that appear to facilitate the study of quantity implicatures. Additional facts related to a full understanding of the phenomenon of quantity implicature interpretations appear to be its strong reliance on the lexicon and EF. Conceptually, syntax must also play a role in these interpretations. Finally, an approach to modeling the relevant variables, such as those we have enumerated, instead of modeling a single multiple regression, has become possible in the domain of ecological studies. This technique could be of use to us in modeling the cognitive processes underlying the processing of pragmatic implicatures. Given these considerations, we are led to the following research questions:

1. Do adults categorically generate scalar implicatures with *algunos*?
2. Do children categorically generate scalar implicatures with *algunos*, and, if so, at what developmental moment do they begin doing so?
3. Can we use the statistical technique of piecewise SEM to hierarchically model the conceptually necessary cognitive components of the psycholinguistic implicature generation system we propose: lexicon, syntax, EF, and the ANS?

7 Methodology

7.1 Participants

In all cases, a parent or guardian signed a university institutional review board-approved informed consent document for children, and adult controls signed their own informed consent documents. Eighty-two monolingual, typically developing Spanish-speaking children in Mexico City took the experiment. Eighteen of them were excluded for not passing our filler items, which we describe below. Most of these children were from the younger part of our sample (12 of them were 4 or 5 years old), leaving us with the 64 whose data we consider. These children's ages ranged between 50 (4;2) and 101 (8;5) months, with a mean age of 77.8 months (7;5) and a standard deviation of 15.2 months. Adult participants ($n = 27$) ranged in age between 218 (18;2) and 434 (36;2) months, with a mean age of 320 months (26;8) and a standard deviation of 55 months. The number of participants per age group is given in Table 1.

7.2 Procedures

7.2.1 Truth value judgment task. We follow the core elements of the original version of the TVJT, from Crain and McKee (1985), though modified by presenting the pragmatic context in our experiments using stop-motion video with digitally recorded audio.

Our TVJT scenarios include plausible dissent, following Crain and McKee, which is the possibility that the answer given at the end of the task, could plausibly be either true or false as a

Table 1. Ages of Participants.

Age	4-year-olds	5-year-olds	6-year-olds	7-year-olds	8-year-olds	Adults	Total
n	7	19	13	14	11	27	91

Figure 3. Scene depicting the third of four children consecutively climbing a ladder to watch television.

function of the context. This is achieved in our scenario by the narrator casting doubt on whether the action under consideration is possible and by having the agents of the action discuss whether they should carry out the action.

15. Los niños están en la casa y quieren subir a ver la tele, pero la escalera es muy alta.
    The children are at home and they want to go upstairs to watch TV, but the ladder is very tall.

After the narrator's comment regarding the height of the ladder, the children huddle to discuss ladder climbing. The narrator then asks the QUD, following Gualmini et al. (2008):

16. ¿Quién va a subir a ver la tele?
    Who is going to go up and watch TV?

The children then consecutively climb the ladder, as in Figure 3, and the task concludes with the narrator producing the critical experimental sentence.

17. Ya sé. Algunos niños subieron la escalera.
    I know. Some children went up the ladder.

This sentence is designed to either be pragmatically felicitous with a quantity implicature, if three of the four children went up the ladder (the fourth one tried, but kept falling off the ladder's lowest rung), or infelicitous if all four of the four children went upstairs, which creates a clash with "some, but not all" quantity implicature.

There are five such experimental scenarios, all with eventive predicates, in which children consecutively carry out a telic action:

18. Algunos niños bajaron la resbaladilla.
    Some children went down the slide.

19. Algunos niños subieron la escalera.
    Some children went up the ladder.

20. Algunos niños brincaron la cerca.
    Some children jumped the fence.

21. Algunos niños cruzaron la calle.
    Some children crossed the street.

22. Algunos niños le dieron vuelta al autobús.
    Some children went around the bus.

Each sentence was paired with both a three out of four (implicature-consistent) video scenario and with a four out of four (implicature-inconsistent) video scenario.

To orient participants to the nature of the task and to make sure that they are paying attention to the task, there are four warm-up items that use quantifiers that children appeared able to use by 4-year old at high levels of proficiency in our pilot work for quantity judgments: *todos* (all) and *ningún* (none). There are two such warm-up sentences with each quantifier. The *todos* (all) sentences are presented in both a four of four context, in which they are true, and in a two out of four context, in which they are false. The *ningún* (none) sentences are presented in a zero out of four context, in which they are true, and in a three of four scenario, in which they are false.

23. Todos los niños pasaron el puente.
    All the children crossed the bridge.

24. Ningún niño pasó el puente.
    No child crossed the bridge.

These sentences are necessary to make sure that participants are paying attention and understand that the task is about quantity. It is for this reason that we include the quantifiers *todos* (all) as well as *ningún* (none), without using the object of our inquiry, which is another member of this scale, *algunos* (some), and its associated implicature. This type of priming is critical to children grasping what aspect of the sentence and the context they are supposed to pay attention to. The warm-up items and our feedback are not sufficient, in and of themselves, to assure that all children are able to succeed at the task if they do not understand the quantity implicature, and it does not train children on the quantifier under study or on any implicature-type inference.

Two additional such sentences are presented as filler items, in the same context–quantifier pairs as in the warm-up items, with the *todos* (all) and *ningún* (none) quantifiers, interspersed with the experimental items (total of four sentences).

25. Todos los niños se metieron en el alberca.
    All the children got in the pool.

26. Ningún niño se metió en el alberca.
    No child got in the pool.

Feedback is given on the warm-up items, but not on the experimental or filler items. Only participants who answered correctly on all four filler items were included in the sample. No adults were excluded for this reason. Out of 82 children, 18 were excluded for this reason, most of whom were from the younger part of our sample (12 of them were 4- or 5-year-olds), leaving us with the 64 whose data we consider. Of the 18 children who were excluded for failing fillers, 15 said “no” when the sentence “No child got in the pool.” was presented. The other three (of a total 82 children) had problems with a sentence containing *todos* (all), which is 3.7% of our total sample.[7]

Scenarios were presented using the SuperLab 5 software (Cedrus Corporation), on 13″ Macintosh MacBook Air laptop computers, in randomized order. Participants indicated acceptance of the final sentence by pushing the “c” key of the keyboard, which has a foam “smiling face” affixed to it and conveyed rejection by pushing the “m” key of the keyboard, which has a foam “sad face” affixed to it. The audio of the experiments was presented through over-the-ear, noise-canceling headphones, attached to the computers. All audios were recorded by a female native speaker of the same variant of Spanish spoken by the children (Mexico City Spanish).

7.2.2 Lexicon. Children’s lexical development was measured using the *Adivinanzas* (“riddles”) subtest of the BELE (Rangel et al., 1988), which is normed in Mexico City. It consists of a series of “clues” as to the nature of the word, as in, “They are what you use to see. You have them on your face and you close them when you sleep. What are they?” In previous work, these results have correlated with the Peabody Picture Vocabulary Test in Spanish (TVIP; Dunn et al., 1986), the number of different words lexical measure, and others, confirming their validity.

7.2.3 Syntax. Our measure of syntax is the argument index (ARG-I). ARG-I is based on the subordination index of Loban (1963), which takes the mean number of clauses per C-unit, where a C-unit consists of an utterance, including at least one clause—no fragments. The subordination index, calculated from Frog Story spontaneous production samples, is based on a story retell procedure, using the “Frog, where are you?” story book (Mayer 1969). The subordination index does not distinguish argument clauses, which appear as part of the argument structure of the matrix clause verb, from adjunct clauses, which occur freely, irrespective of the verb’s argument structure. In the work of Grinstead et al. (2020), we show that the ARG-I, the mean number of argument subordinate clauses over C-units, correlates with MLUw (from independent transcriptions) expressive morphosyntax measure, the elicited production expressive morphosyntax measure of the BELE, and with the grammatical comprehension receptive morphosyntax measure of the BELE. Following Castilla-Earls et al. (2015), we used an iterative process to check the accuracy of transcripts, C-unit segmentation, and coding for ARG-I. Following the initial transcription and coding, a second examiner independently viewed each frog story transcription, while examining the transcript for errors in transcription, segmentation, and coding. The original and second examiner then discussed discrepancies until they reached agreement. The transcriptions were created following the CHAT conventions of the Child Language Data Exchange System (CHILDES) project (MacWhinney, 2000).

7.2.4 Executive function. We use three tasks from the EXAMINER Battery (Kramer et al., 2014), including the flanker task of inhibition (Eriksen & Eriksen, 1974), the set-shifting task of attention (Kray & Lindenberger, 2000), and the dot counting task for auditory working memory (Case et al., 1982), following the Miyake et al. (2000) model of EF. The flanker and set-shifting tasks produce a score that is a regression coefficient of the total number correct by reaction time. Dot counting is the highest number of items that could be remembered.

7.2.5 The approximate number system. We used the Siegler and Opfer (2003) non-symbolic number line task as a measure of ANS estimation. In this task, children estimate the position of a set of green dots on a number line that goes from no green dots, at the left edge of the line, to 30 green dots, at the right edge of the line. The task requires that a number of calculations (referred to as "mensuration") be performed on children's visual ANS representation. Children have been shown to traverse a developmental sequence that begins with them producing log estimates of the quantities they are presented with, consistent with Fechner's Law, and culminating in adult-like linear estimates of the quantities they are presented with. This trajectory is referred to as the *log-to-linear Shift*. Siegler and Opfer (2003) calculate a proportion of estimates that are log in character, which they refer to as Lambda, running from 0, which is 100% linear, to 1, which is 100% logarithmic.

8 Results

8.1 Descriptive statistics

8.1.1 Categorical judgments and algunos. With regard to our outcome variable, the TVJT, the results are as in Figure 4.

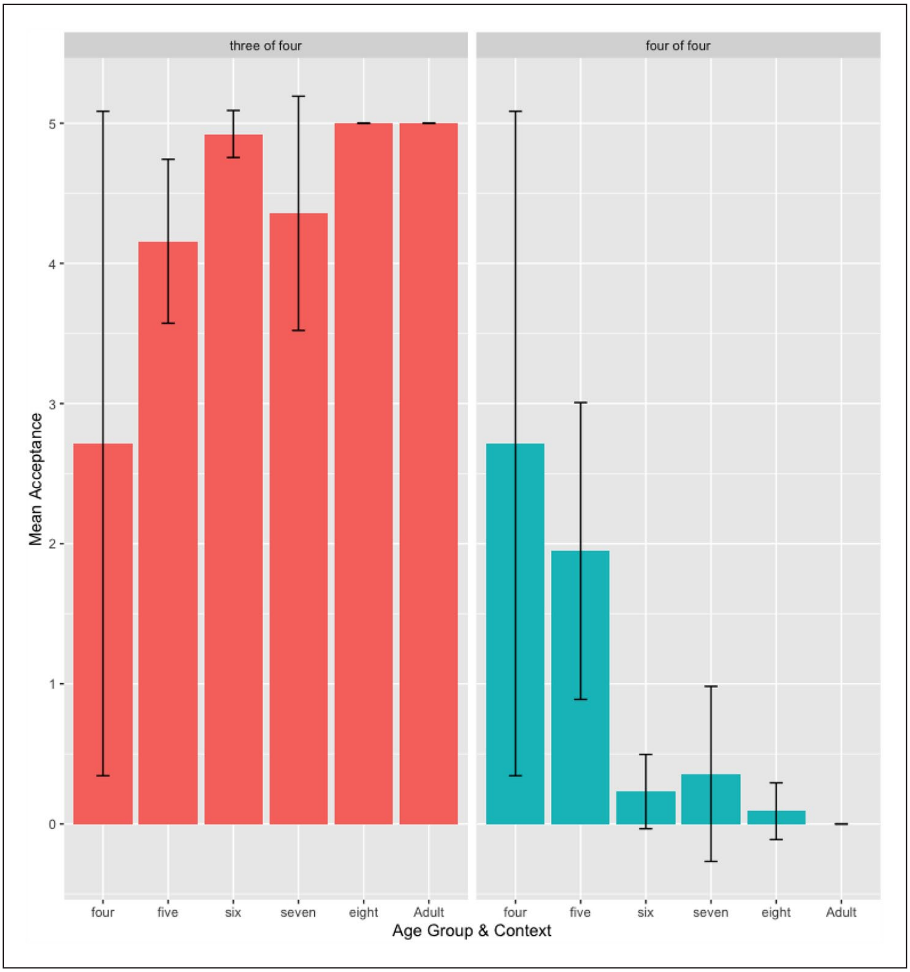

Figure 4. Acceptance of *algunos* in implicature-consistent (three out of four) and in implicature-inconsistent (four out of four) contexts.

In our descriptive statistics, we find a number of prominent patterns. To begin with, adults indeed generate implicatures categorically. That is, on the right side of Figure 4, which has blue bars, the bar representing the number of times adults accepted *algunos* in *todos* contexts is an average of 0 (0/135). This pattern emerged in our piloting of this version of the TVJT and persisted through all 27 adult controls, on all experimental items, providing us with some confidence as to how adults interpret our stimuli. For children, our 6-, 7-, and 8-year-old groups show behavior very similar to that of adults, with average acceptance rates of *algunos* in *todos* contexts, of far less than one scenario out of five (.23/5, .36/5, and .09/5, respectively). In contrast, 4- and 5-year-old children show mean rates of acceptance that are higher (2.71/5 and 1.95/5, respectively). These mean rates, however, are somewhat deceptive, in the sense that one could conclude that they imply that individual children were primarily accepting two or three scenarios with *algunos* in *todos* contexts. In reality, children in the 4- and 5-year-old age bands show a bimodal distribution. In the implicature-inconsistent condition, of the seven children in the 4-year-old group, three of them categorically generate implicatures (acceptance = 0), while four of them categorically fail to generate implicatures (acceptance = 4 or 5). These three categorical implicature generators could be seen as evidence that, chronologically at least, there are 4-year-olds who categorically generate implicatures.

In the 5-year-old group, there is also a type of bimodality, with the largest group (9/19) categorically generating implicatures (acceptance = 0), with five children never generating implicatures (acceptance = 5) and five more children in the middle (accepting one, two, three, or four implicature-generating scenarios), illustrated in Figures 5 and 6.

The 5-year-old group ($n = 19$) is the only group with children who give evidence of having a statistical "middle" (acceptance of two or three scenarios). It is interesting to note that the three children who give "middle" answers in the implicature-inconsistent condition, illustrated in Figure 6, are not the same three children who do so in the implicature-consistent condition, with one exception. This situation of giving inconsistent answers could be the result of them having low degrees of confidence in their answers, perhaps because learning is taking place. They are a small percentage of the 5-year-old group (3/19 or 16%).

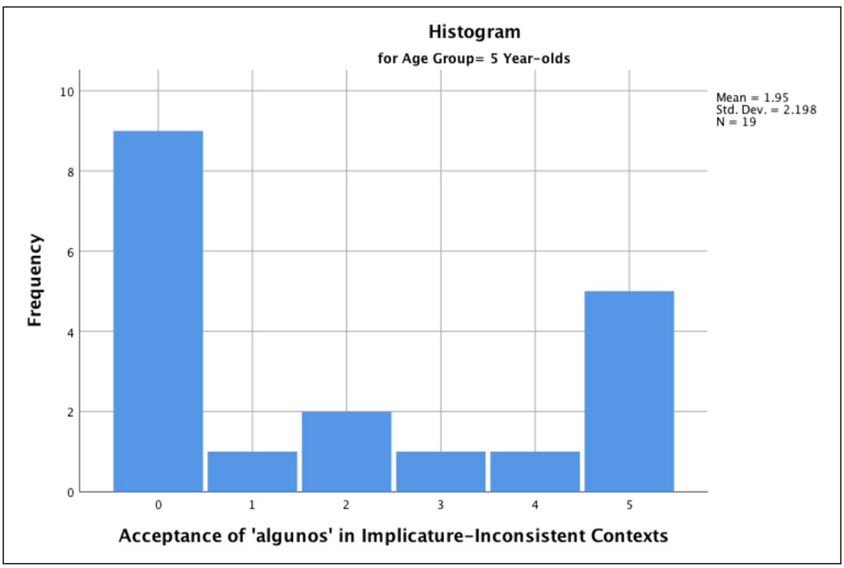

Figure 5. Histograms of 5-year-old children's acceptance of *algunos* in implicature-inconsistent (four out of four) contexts.

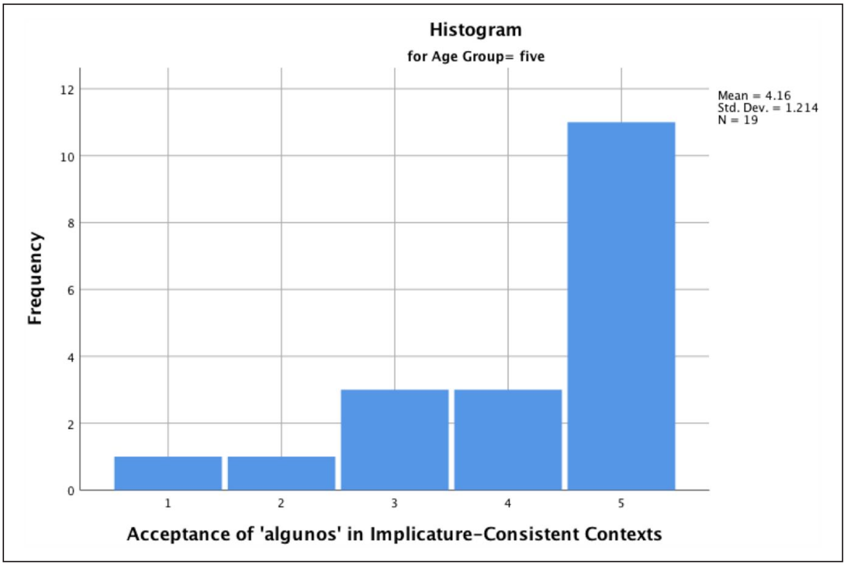

Figure 6. Histograms of 5-year-old children's acceptance of *algunos* in implicature-consistent (three out of four) contexts.

The second pattern we observe is that a small number of children in the sample (the non-adult-like extreme of the bimodal distribution) appear to have exactly the opposite of adult judgments of the contextual interpretations of *algunos*, and consistently so. Specifically, in the implicature-consistent "3 out of 4" context, there are four children, three in the 4-year-old group and one in the 7-year-old group, who consistently reject *algunos*. These same four children consistently accept *algunos* in the implicature-inconsistent context in our experiment. It is a small group, obviously (4/64 or 6%), but their behavior does not appear random, but rather appears to be systematically the semantically polar opposite of what adults do. To be clear, we are merely describing our data and nothing theoretical hangs on this finding, but the observation may be of interest in future studies of the phenomenon.

8.2 Summary of outcome and predictor variable values across age groups

The means and standard deviations of outcome variable and predictor variables, across our sample's age bands, are summarized in Table 2.

8.3 Inferential statistics

8.3.1 Pearson's correlations. In Table 3, we see Pearson's correlations of the study variables. Note that in Tables 2 and 3, greater implicature generation is indicated by lower acceptance of incongruous quantifier–context pairing (*algunos* in *todos* contexts), which means that variables that correlate with implicature generation will produce negative correlation coefficients.

8.3.2 Age comparisons. To understand when age bands of children are indistinct from adults, we compare each age band to each other, and most crucially, to adults. The intuitive differences are illustrated in Figure 4. Given the nature of the count distribution into which the data fall, we used

Table 2. Means and Standard Deviations of Study Variables by Age Group.

Years of age	TVJT (implicature-consistent—three of four)	TVJT (implicature-inconsistent—four of four)	Lexicon (*Adivinanzas*)	Syntax (ARG-I)	ANS (non-symbolic number line)	Inhibition (flanker)	Attention (set-shifting)	Working memory (dot counting)
4	2.71 (2.56)	2.71 (2.56)	11.57 (5.32)	.04 (.03)	.71 (.49)	4.79 (1.07)	4.43 (.79)	5.57 (1.51)
5	4.16 (1.21)	1.95 (2.20)	14.53 (5.86)	.03 (.02)	.86 (.29)	5.51 (.86)	5.68 (1.92)	6.79 (2.12)
6	4.92 (.28)	.23 (.44)	15.38 (3.59)	.06 (.03)	.70 (.42)	5.89 (.48)	5.38 (2.22)	10.08 (4.07)
7	4.36 (1.45)	.36 (1.08)	19.57 (3.55)	.04 (.02)	.56 (.48)	5.85 (.51)	6.07 (2.37)	10.50 (4.20)
8	5.00 (0)	.09 (.30)	21.82 (2.93)	.05 (.03)	.59 (.49)	6.11 (1.15)	6.18 (1.54)	11.73 (3.88)
Adults	5.00 (0)	0 (0)	–	–	–	8.16 (.72)	7.38 (.86)	14.30 (4.10)

TVJT: truth value judgment task; ARG-I: argument index; ANS: approximate number system.

Table 3. Pearson Product Moment Correlations of Study Variables.

	Algunos	Lexicon	Syntax	ANS	Inhibition	Attention	Working Memory
Algunos		–.446***	–.244	.126	–.432***	–.430***	–.362**
Lexicon			.072	–.296*	.425***	.382**	.329**
Syntax				.023	.073	.208	.299*
ANS					–.151	–.050	–.234
Inhibition						.691***	.452***
Attention							.317*

ANS: approximate number system.
*$p < .05$, **$p < .01$, ***$p < .001$.

a Poisson regression to model it. The equivalent of an analysis of variance (ANOVA) post hoc test for Poisson regression (the estimated marginal means test) shows adults as significantly different from 4-, 5-, and 6-year-olds. With respect to the 6-year-olds, this difference stems from the log-odds of count probability calculation in the Poisson regression with zero variance in adult's 100% implicature judgments. For an intuitive comparison of age group mean implicature, compare 4-year-olds (M=2.71, SD=2.56) and 5-year-olds (M=1.95, SD=2.20), who appear quite different from 6-year-olds (M=.23, SD=.44), who seem more similar to 7-year-olds (M=.36, SD=1.08), 8-year-olds (M=.09, SD=.30), and adults (M=.0, SD=.0). In fact, if we model these data as a continuous variable, using a one-way ANOVA, as in the work of Guasti et al. (2005), 6-year-olds are not statistically different from adults, by post hoc test.[8] This zero variance problem does not arise for our predictive multiple regression models in the remainder of the paper because only child data are considered.

8.3.3 Poisson regression with syntax and lexicon. Our primary dependent variable, implicature generation has a count distribution in our sample, which is most appropriately modeled in generalized linear models using the Poisson regression, with a log-odds linking function. This linking function is particularly appropriate, given the bimodality component of the distribution we have discussed.

Because our model of language critically assumes that syntax and lexicon must combine to produce semantic interpretations, we begin with a multiple Poisson regression that includes both syntax (ARG-I) and lexicon (*Adivinanzas*). With mean acceptance of *algunos* in implicature-inconsistent contexts as the outcome variable, both syntax (ARG-I) and lexicon (*Adivinanzas*) are significant predictors, accounting for unique variance. The likelihood ratio chi-square test (–91.15) indicates that the full model was a significant improvement over a null (no predictors) model ($p < .001$), and the Akaike information criterion (AIC), which is used to compare regression models by balancing the number of parameters included with the amount of variance accounted for, was 188.30. For every one-unit increase in ARG-I, the log count of *algunos* acceptance decreased (increased in implicature generation) by 21.26 (B=–21.26, SE=6.43, p=.001) and for every one-unit increase in *Advinanzas* score, the log count of *algunos* acceptance decreased by .123 (B=–.12, SE=.02, $p < .001$), as in Table 4.

8.3.4 Executive function. Our initial Pearson's correlations show that all of the EF variables correlated with *algunos* implicature generation. Given the conceptual role played by EF in choosing words in the lexicon and in matching sentences to pragmatic contexts, the next logical step in constructing our model is to include our EF variables of inhibition, auditory working memory and attention, but how many and which ones? As illustrated in Table 4, Poisson regression models that

Table 4. Regressions of Syntax, Lexicon, and EF Variables on *algunos* in Implicature-Inconsistent Contexts. Coefficients (Standard Errors) and Standardized Coefficients.

Predictors	Null	Model 1	Model 2	Model 3	Model 4	Model 5
Syntax		–21.257*** (6.427) *–.542*	–15.993* (6.646) *–.407*	–21.225*** (6.558) *–.541*	–23.413*** (6.826) *–.597*	–19.271** (7.241) *–.503*
Lexicon		–.123*** (.021) *–.673*	–.096*** (.023) *–.525*	–.089*** (.023) *–.489*	–.089*** (.022) *–.489*	–.073** (.024) *–.399*
Working memory			–.111* (.046) *–.437*			–.066 (.052) *–.261*
Attention				–.438** (.139) *–.385*		–.217 (.170) *–.190*
Inhibition					–.276*** (.072) *–.452*	–.153 (.103) *–.251*
Constant	.015 (.124)	2.543*** (.358) *–.374*	2.816*** (.372) *–.430*	4.402*** (.675) *–.444*	3.551*** (.450) *–.494*	4.216*** (.682) *–.517*
AIC	232.08	188.30	184.12	180.52	177.06	178.20

AIC: Akaike's information criterion.
$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

include syntax, lexicon, and each of the EF variables, individually, are significant. A model that includes syntax, lexicon, and all of the EF variables, together, however, shows none of the EF variables to be significant, presumably because of collinearity, which has been widely reported (e.g., Miyake et al., 2000). The respective AIC values for each of the syntax + lexicon + EF models are given in Table 4, which also gives standardized coefficient values to make the relative force of each variable more intuitively interpretable.

For the lowest (and therefore the most explanatory) AIC model, with inhibition included, the log likelihood ratio chi-square test (–84.53) shows the model to be a significantly better fit than the null model ($p < .001$). On this model, for every one-unit increase in ARG-I, the log count of *algunos* acceptance decreased by 23.413 ($B = –23.413$, $SE = 6.825$, $p = .001$), for every one-unit increase in *Adivinanzas* score, the log count of *algunos* acceptance decreased by .089 ($B = –.089$, $SE = .022$, $p < .001$) and for every one-unit increase in the flanker (inhibition) score, the log count of *algunos* acceptance decreased by .276 ($B = –.276$, $SE = .072$, $p < .001$). The standardized coefficients show syntax, lexicon and inhibition have similar effects on the outcome.

8.3.5 Shortcomings of an unstructured multiple regression. In the significant model we have just reported, schematized in Figure 7, it was possible to make the most straightforward claim about scalar implicature generation associated with the existential quantifier *algunos* in Spanish, namely, that syntax must concatenate lexical items into sentences, which inhibition must then exclude as appropriate descriptions of the infelicitous pragmatic contexts in which they were presented.

However, we are unable, in this model, to account for the role that the existing literature claims that inhibition also plays in choosing lexical items from the lexicon, before syntax concatenates them (see Gangopadhyay et al., 2019; McMurray et al., 2019, inter alia). Furthermore, though our

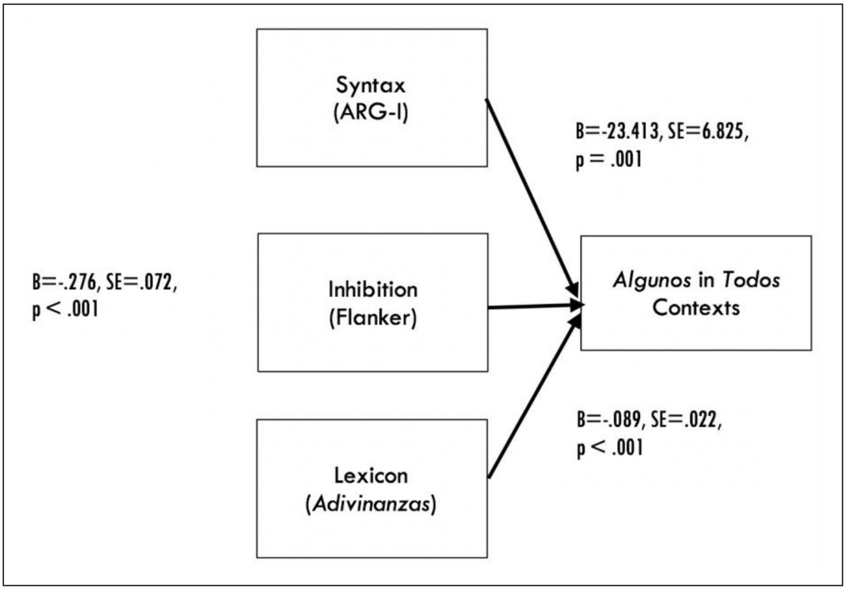

Figure 7. Syntax, lexicon, and inhibition predict *algunos* interpretations in contexts where *todos* ("all") would be correct, multiple Poisson regression—AIC = 177.06.

ANS variable correlated with our lexical measure, above, it does not correlate with, nor does it significantly predict implicature generation in a regression model ($p > .05$). Nonetheless, our lexical refraction hypothesis proposes that ANS provides non-discrete quantity knowledge to the lexicon so that quantity-sensitive words, including natural language quantifiers, such as the existential *algunos*, can come to convey the quantal, discrete meaning "more than exactly one." It is conceptually possible, then, that ANS plays a role in quantifier interpretation, but at the sub-lexical level. However, there is no statistical means of testing such a relationship in a flat-structured multiple regression. For these reasons, we now turn to a piecewise SEM of our data.

8.3.6 Piecewise SEM. In Figure 8, we fit a piecewise SEM to our five variables. Following Shipley (2000, 2002, 2009) using the PiecewiseSEM package in R (Lefcheck, 2019), we find a *C* statistic of 2.783, with a *p*-value of .595 and on 4 degrees of freedom, based on the two locally estimated multiple regression equations that make up this SEM. We note, again, that a *p*-value of above .05 is desirable for attribution of causality in the sense that we are comparing the model-wide *C* statistic to the chi-square distribution, to determine whether it is likely to have occurred by chance, which would be the case if the *p*-value were less than .05. That is, we want to establish that an alternative causative model using the variables we have included in this model is not significantly explanatory. The *p*-value of above –.05 tells us that just this outcome has obtained. The rightmost equation, predicting *algunos*, is a Poisson multiple regression, while the model predicting lexicon is an ordinary least squares (OLS) multiple regression. Fisher's *C* statistic indicates that the model shows good fit.

Furthermore, in Table 5, we can see in the tests of directed separation that the two other possible paths, given our endogenous variables, both produce non-significant results (e.g., the *p*-value of over .05 associated with the *C* statistic derives from these values), supporting our hypothesis that just the path indicated among the exogenous and endogenous variables considered is a close fit to the data.

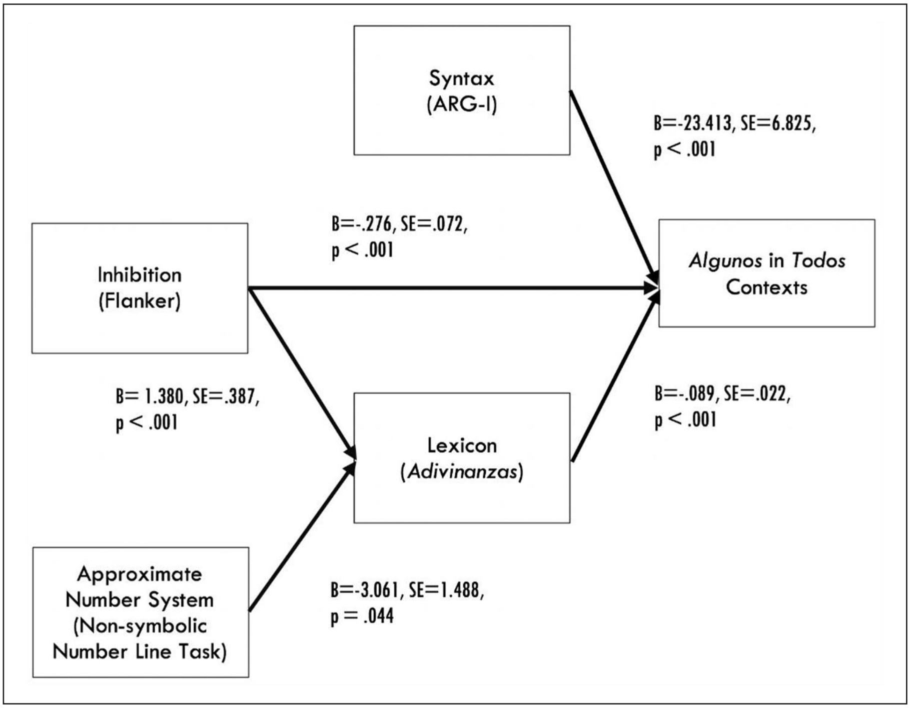

Figure 8. Syntax, lexicon, and inhibition predict *algunos* interpretations in contexts where *todos* ("all") would be correct; inhibition and approximate number predict lexicon—Fisher's $C = 2.783$, $p = .595$, $df = 4$, AIC = 18.783. A data file (.csv) and an R script to reproduce the results in Figure 8 are available here: https://osf.io/57vxr/.

Table 5. Non-significant Tests of Directed Separation.

Independence claim	*df*	Critical value	*p*-value
Lexicon ~ syntax + inhibition + ANS	58	.680	.499
Implicature ~ ANS + syntax + inhibition + lexicon	57	–.677	.498

ANS: approximate number system.

Other possible combinations of variables could include our working memory variable or our attention variable in place of inhibition. With working memory in place of inhibition, Fisher's *C* statistic = 1.387, $p = .846$, but number is no longer predictive of lexicon, suggesting that working memory plays a role in our number measure. The AIC for the model = 17.387. With attention in place of inhibition, the Fisher's *C* statistic = .549, $p = .969$, indicating good fit. The model's AIC = 16.549. Thus, both the attention and the inhibition models show high goodness-of-fit and show significant local relationships among the variables. The working memory model falls short in that number and working memory are too collinear to allow good fit. All three SEMs, it should be said, appear similar, which is consistent with the widely replicated finding that though the components of EF are independent, they are also somewhat correlated. However, the models with attention and inhibition appear statistically to be the best descriptions of at least our measures of these constructs.

Table 6. Power Comparison (n = 64) of Piecewise (Locally Estimated) and Traditional (Globally Estimated) SEM of the Relationships Depicted in Figure 8 Predicting Implicature Generation with *algunos*.

	Locally estimated		Globally estimated
	Lexicon regression (inhibition, ANS)	Implicature regression (syntax, lexicon, and inhibition)	Entire model
Power with $n = 64$	.98	.99	.21

ANS: approximate number system.

Given our earlier comments regarding the advantage of piecewise SEMs over traditional SEM models with respect to power, it is worthwhile illustrating the difference. Because our SEM is mixed between a linear model and a generalized linear model, estimating power is not entirely straightforward. However, if we fit both a traditional SEM and a piecewise SEM with only linear regressions, power can be easily estimated, as in Table 6.

As should be clear from Table 6, the power in our piecewise SEM is quite strong, while the power associated with the traditional model is inadequate for making claims regarding the system of variables, which we have proposed to play a role in children's quantity implicature interpretations. In particular, Cohen (1988) argues that statistical power should be at least .8, meaning that our globally estimated SEM is woefully underpowered, while the piecewise SEM is amply powered in each of the equations. For researchers who have carried out such multi-task protocols, and who are interested in modeling cognitive systems, it should be clear that such an objective is achievable without the participation of four or five times the number of children that participated here, to appropriately power a traditional, globally estimated SEM.

9 Discussion

Returning to our research questions, we see first that adults indeed categorically generate "some, but not all" quantity implicatures in our experiment. As we have pointed out, this contrasts with much of the existing literature in which adults not only fail to generate implicatures 100% of the time, as in this study, but sometimes generate implicatures at rates as low as 59% of the time. We have a relatively high level of confidence that adults understood the task that they were being asked to perform and the results appear categorical. We attribute these high levels of adult implicature interpretations to having omitted the confounds we have reviewed and to having included experimental properties that promote implicature generation, as discussed.

Furthermore, children appear to generate implicatures categorically from our youngest participants. Three of the seven 4-years-olds and nine of the 19 5-year-olds never accepted *algunos* in implicature-inconsistent contexts, exactly as adults. That is, roughly half (12/26, 46%) of the 4- and 5-year-olds are adult-like. In terms of development, it appears possible for preschool-aged children to generate the quantity implicature.

The 5-year-old group may illustrate the transition to adult-like implicature understanding. As noted, nine of the 19 children in the 5-year-old group categorically generated implicatures. However, the remaining 5-year-olds appeared to undergo an implicature learning process, during which those who are categorical in matching *algunos* with implicature-consistent contexts did not do so with implicature-inconsistent contexts, and vice versa.

Beginning at age 6, children categorically generate implicatures, with very little variance within the age band. This is younger than the earliest TVJT results that used covert partitives, namely, Guasti et al. (2005), who showed that fifteen 7-year-old Italian-speaking children would generate implicature

75% of the time, compared to adults' 83% of the time, which was not significantly different. Our 4- and 5-year-olds were not different from each other, but were different from 6-, 7-, and 8-year-olds, and adults' 100% implicature generation, as well. Our results are similar to those of Miller et al. (2005), who showed with an act-out task using *algunos* that child Spanish speakers could react appropriately to the "some, but not all" interpretation at 88%–97% implicature generation levels, with adults generating implicatures 100%. Taken together, these findings show, with both a more expressive (act-out) and a more receptive (TVJT) measure, that child Spanish speakers are capable of drawing quantity implicature interpretations in adult-like ways by 4 years old and do so uniformly by 6 years of age.

A piecewise SEM allowed us to statistically show significant multiple regression relationships predicting lexicon as an outcome variable with number and one of the two the EF components of either inhibition or attention, as predictors. The first predictor variable, from EF, is plausibly related to attending selectively to, or to inhibiting, the set of possible quantifiers that could have been chosen among the competing quantifiers that are linked together on the quantity scale {todos, la mayoría, muchos, algunos, pocos . . .} in the lexicon. The second predictor variable, number, plausibly represents a part of the mind that is not inherently linguistic and is presumably shared by a wide array of non-human animals (e.g., Meck & Church, 1983 and studies reviewed in the work of Gallistel, 1990). The ANS representations provided, make a plural existential quantifier meaningful, namely, "more than exactly one." This is consistent with our lexical refraction hypothesis, which makes the same claim with regard to the cardinal numbers in the count routine, plural marking agreement morphemes that are bound to determiners, adjectives, and verbs in Spanish, along with other bound and free morphemes in the lexicon that express quantity information.

Lexicon, in turn, is shown to significantly predict *algunos* implicatures, together with syntax and a second use of inhibition or attention. This is consistent with other work (Pratt et al., 2020) showing that four distinct lexical measures are capable of not only significantly predicting *algunos* implicature generation by themselves, or in a composite, but that together they identify implicature generators and non-generators with high levels of sensitivity (91%) and specificity (100%). Similarly, it was shown in the work of Grinstead et al. (2019) that all three EF variables were predictive of *algunos* interpretations in implicature-inconsistent contexts.

Syntax is also significantly predictive of *algunos* implicatures in this model. Showing a relationship between syntax and implicature generation, in which syntax is not collinear with lexicon, is in fact difficult. This is so because many measures of syntax, such as MLU, or many standardized measures, incorporate a high degree of lexical knowledge in them, which makes entirely disentangling lexicon from syntax challenging. The ARG-I, however, appears to be statistically quite independent of our lexical measure and may be an effective means of measuring phrasal syntactic competence. It accounts for unique variance in the implicature interpretation associated with *algunos*. We take these results to be consistent with an interpretation of Reinhart (1998, 2004) to the effect that syntax must work harder, essentially, when an interface with discourse-pragmatics is called for by the construction be produced or interpreted.

9.1 SEMs and the parallel architecture

In our project, we have attempted to model the interaction of multiple domains of cognition in a system that explains how at least some of the relevant constructs conspire to produce conversational implicatures. Thinking about how this many domains of cognition interact is not a frequently addressed topic in cognitive science, though there are exceptions. One explicit model that includes many of the relationships we are concerned with is the parallel architecture of Jackendoff (2002). Of greatest relevance in this model, Jackendoff proposes that each sub-domain of language (syntax, phonology, semantics, etc.) has its own processing mechanisms and he points to working memory, specifically.

In our study, we are only able to model a subset of the variables Jackendoff considers, namely, one domain-specific, non-linguistic cognitive domain in the form of number, three domain-general processing-related domains of cognition, in the form of the three EF variables, and three linguistic sub-domains, including lexicon, syntax, and semantics–pragmatics, which would be referred to as conceptual structure, in Jackendoff's terms. Our model is consistent with Jackendoff's claim that processing applies to multiple domains of linguistic cognition, in that inhibition applies both to individual lexical items and to the semantic–pragmatic interpretations of full sentences. We will, though, have less to say about working memory, which fit less well than inhibition with our model, at least as we measured it, with respect to the particular construction under consideration.

We would not want rule out an important role in the process for working memory, given its usual activity in processes that also involve inhibition. We also take our lexical refraction model of number and lexicon to be consistent with the spirit of Jackendoff's proposal, which explicitly considers the interaction of language and spatial cognition (e.g., Landau & Jackendoff, 1993). Our model is more concerned with numerical cognition and its interaction with language, because we are focused on quantifiers, but there are similar considerations, in the sense that lexical items and numerical representations overlap in certain ways in what they represent, while not being identical in the information they convey (as in, e.g., Grinstead et al., 1998).

9.2 Piecewise SEM for the social and behavioral sciences

While we hope to have made a contribution to what is known descriptively and methodologically with respect to children's pragmatic implicature interpretations and theoretically to what is thought about the cognitive systems that produce them, we also believe that it could be fruitful for cognitive scientists concerned with the interfaces among mental faculties to adopt piecewise SEM. Perhaps the greatest obstacle to this type of systemic theoretical work has long been the statistical power required to make traditional SEMs plausible. The only logical remedies to this obstacle have been to either reduce the complexity of the systems modeled, and only look at simple, local relationships; or to study only the most conventional of complex cognitive relationships that are very likely to show good fit, but unlikely to tell us much we do not already know about cognition. After all, one cannot know a priori that the relationships explored will in fact be significantly related, as measured. To embark on such a protocol with hundreds of children involves substantial risk that one's model, though insightful, might not show the statistical properties necessary to convince readers. In contrast, if we are, for example, able to model the interplay of five distinct cognitive variables in the construction of pragmatic implicature interpretations, with good fit and high power, in a sample of only 64 children, then it may prove easier for cognitive scientists to test less probable or conventional hypotheses, which may nonetheless be correct. Surely, this type of increased freedom to model cognitive systems with smaller participant samples holds out the promise of a greater diversity of ideas to be considered and to precipitate a commensurate increase in scientific insight. We aspire to conduct future work that will help disentangle the relative contributions of each domain in production, as well, and in other semantic–pragmatic construction types.

Acknowledgements

The authors are grateful to Craige Roberts, John Opfer, Ramón Padilla-Reyes, Melissa Nieves-Rivera, Jhovana Sandoval-Estrada, Michell Zúñiga-Espinosa, Alejandro López-Ramos, Debi Gómez, Olivia Mancera-Unikel, Eva Sifuentes-Monroy, Pedro Ruiz-Curcó, and Mary Carmen Gasca for the help in carrying out the protocol. Le dedicamos el artículo a nuestro gran amigo y maestro, el Profesor Javier Gutiérrez-Rexach, quien nos inspiró y nos abrió las puertas a los indefinidos y a tantos aspectos más de la semántica.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project was funded by a grant from the National Science Foundation (NSF-1551903).

ORCID iD

John Grinstead https://orcid.org/0000-0001-6667-8250

Notes

1. Pouscoulous et al. (2017) refer to this obstacle as "metalinguistic" in character in their follow-up experiment, which overcomes this obstacle.
2. Of course, it is possible that children do not yet have adult-like knowledge of the uniqueness presupposition at a given moment in their semantic development, which adds an additional variable to be controlled in overt partitive construction experiments and potentially muddles any finding other than a demonstration of adult-like knowledge. That is, if children are not performing in an adult-like way, how do we know that this is not attributable to problems with their grasp of the uniqueness presupposition? Although we suspect that this is not a problem by 6 years of age, see Maratsos (1974, 1976) and much subsequent research on younger children's interpretations of definite articles, which concludes that English-speaking 4-year-olds are not 100% adult-like in their definite interpretations. In either case, an additional variable is included in such experiments, though it has not ever been mentioned, much less controlled, to our knowledge.
3. MLUw from unstructured spontaneous production and age significantly correlated ($r = .441$, $p < .001$, $n = 64$) as did MLUw from frog story retell and age ($r = .471$, $p < .001$, $n = 64$).
4. "Domain-general inhibition" is not an atheoretic term, inasmuch as it presupposes the existence of both a domain-general and a domain-specific version of inhibition. It is also, however, not without empirical substantiation. Blomquist (2017) and Blomquist and McMurray (2018) show that a lexicon-internal type of phonological inhibition is predictive of reading development, but that Stroop task scores, a putative measure of domain-general inhibition, are not predictive, consistent with the hypothesis that these two types of inhibition may be independent of one another.
5. A recent paper (Shipley & Douma, 2020) has generalized the χ^2 statistic to locally estimated models that is directly comparable to the one obtained from global estimation methods, and we refer the reader to this reference for further information.
6. Although progress is being made on this front (see Shipley and Douma, 2021).
7. Two of the *todos* (all) filler-failure children were 7 years old and one was 5 years old, which makes this seem unlikely to be a developmental effect. The children in the "No child got in the pool" video put their hands in the pool to see if it is cold. This may have been construed as "getting in the pool." The distribution of the children removed from the sample is quite different and non-adult-like, compared to their age-mates, which suggests that they either did not understand the task, were not paying attention or were misled by the "getting in the pool" video. In any of these cases, their removal from the sample is warranted and makes the generalizations less noisy.
8. To do this, we fit a one-way ANOVA with acceptance of *algunos* in implicature-generating contexts as the dependent variable, with six values for the age variable, each corresponding to an age category: 4-, 5-, 6-, 7-, 8-year-olds, and adults. Results show a significant main effect, $F(5, 83) = 9.194$, $p < .001$, with an adjusted r^2 of .315, and a partial eta-squared effect size measure of .354. Post hoc Tukey's multiple comparisons test shows that 4- and 5-year-olds are not different from one another ($p > .05$), but both 4- and 5-year-olds are each significantly different from all other groups ($p < .05$). Critically, adults were significantly different from 4- and 5-year-olds, but were not different from 6-, 7-, or 8-year-olds, as illustrated in Figure 4.

References

Agrillo, C., Piffer, L., Bisazza, A., & Butterworth, B. (2012). Evidence for two numerical systems that are similar in humans and guppies. *PLoS One*, *7*(2). https://doi.org/http://dx.doi.org.proxy.lib.ohio-state.edu/10.1371/journal.pone.0031923

Antoniou, K., Katsos, N., & Cummins, C. (2016). Why only some adults reject under-informative utterances. *Journal of Pragmatics*, *99*, 78–95.

Astruc, L., Payne, E., Post, B., Vanrell, M. D. M., & Prieto, P. (2012). Tonal targets in early child English, Spanish, and Catalan. *Language and Speech*, *56*(2), 229–253.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*, 83–100.

Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, *3*, 15–70.

Blomquist, C. (2017). *War of the words: Development of inter-lexical inhibition in typical children* [Undergraduate Honors thesis]. University of Iowa.

Blomquist, C., & McMurray, B. (2018, November). *The development of inter-lexical inhibition from 1st grade to adulthood* [Paper presentation]. Boston University Conference on Language Development, Boston, MA, United States.

Carlson, G. (1977). *Reference to kinds in English* [Doctoral dissertation]. MIT.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*, 386–404.

Castilla-Earls, A., Petersen, D., Spencer, T., & Hammer, K. (2015). Narrative development in monolingual Spanish-speaking preschool children. *Early Education and Development*, *5*(8), 1166–1186.

Chierchia, G., Crain, S., Guasti, M. T., & Thornton, R. (1998). "Some" and "Or": A study on the emergence of logical form. *Proceedings of the Annual Boston University Conference on Language Development*, *22*(1), 97–108.

Clifton, C., & Frazier, L. (2012). Discourse integration guided by the "Question under Discussion." *Cognitive Psychology*, *65*, 352–379.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum.

Collingwood, R. G. (1940). *An essay on metaphysics*. Oxford University Press.

Crain, S., & McKee, C. (1985). The acquisition of structural restrictions on anaphora. In S. Berman, J. Choe, & J. McDonough (Eds.), *Proceedings of NELS 15* (pp. 94–110). GLSA.

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology*, *54*(2), 128–133.

Dehaene, S. (1997). *The number sense*. Oxford University Press.

Dunn, L. M., Lugo, D. E., Padilla, E. R., & Dunn, L. M. (1986). *Test de Vocabulario en Imágenes Peabody* [Peabody Picture Vocabulary Test]. American Guidance Services.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon identification of a target letter in a non-search task. *Perception and Psychophysics*, *16*, 143–149.

Feeney, A., Scrafton, S., Duckworth, A., & Handley, S. (2004). The story of some: Everyday pragmatic inference by children and adults. *Canadian Journal of Experimental Psychology*, *58*(2), 121–132.

Feigenson, L., Dehaene, S., & Spelke, E. S. (2004). Core systems of number. *Trends in Cognitive Science*, *8*, 307–314.

Gallistel, C. R. (1990). *The organization of learning*. MIT Press.

Gangopadhyay, I., Ellis Weismer, S., & Kaushanskaya, M. (2019). Domain-general inhibition and lexical processing in monolingual T and bilingual children: A longitudinal approach. *Cognitive Development*, *49*, 68–80.

Grace, J. B., Scheiner, S. M., & Schoolmaster, D. R. (2015). Structural equation modeling: Building and evaluating causal models. In G. A. Fox, S. Negrete-Yankelevich, & V. J. Sosa (Eds.), *Ecological statistics: Contemporary theory and application* (pp. 168–199). Oxford University Press.

Grinstead, J., MacSwan, J., Curtiss, S., & Gelman, R. (1998). The independence of language and number. *Proceedings of the Annual Boston University Conference on Language Development*, *22*(1), 303–313.

Grinstead, J., Padilla-Reyes, R., & Flores-Avalos, B. (2019, November). *Inhibition, general lexical development and the quantity implicature in child Spanish* [Paper presentation]. Boston University Conference on Language Development, Boston, MA, United States.

Grinstead, J., Padilla-Reyes, R., & Nieves-Rivera, M. (2021). A collective-distributive Pragmatic scale and the developing lexicon. *Language Learning and Development*, *17*(1), 48–66.

Grinstead, J., Sandoval-Estrada, J., & Zúñiga-Espinosa, M. (2020). *Working memory and subordination indices* [Paper presentation]. Speer Lab Colloquium, The Ohio State University, Columbus, OH, United States, October.

Grinstead, J., Thorward, J., Ross, S. M., & Maynell, L. (2010, November). *Vowel reduction, pitch accent and scalar implicatures in child English* [Paper presentation]. Boston University Conference on Language Development, Boston, MA, United States.

Groenendijk, J., & Stokhof, M. (1984). *Studies on the semantics of questions and the pragmatics of answers* [Doctoral dissertation]. University of Amsterdam.

Gualmini, A., Hulsey, S., Hacquard, V., & Fox, D. (2008). The Question-Answer Requirement for scope assignment. *Natural Language Semantics*, *16*, 205–237.

Guasti, M. T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes*, *20*(5), 667–696.

Gutierrez-Rexach, J. (2001). The semantics of Spanish plural existential determiners and the dynamics of judgment types. *Probus*, *13*(1), 113–154.

Gutierrez-Rexach, J. (2010). Varieties of indefinites in Spanish. *Language and Linguistics Compass*, 680–693.

Houston, D., Chen C-h Monroy, C., & Castellanos, I. (2020). How early auditory experience affects children's ability to learn spoken words. In M. Marschark & H. Knoors (Eds.), *The Oxford handbook of deaf studies in learning and cognition* (pp. 123–137). Oxford University Press.

Huang, Y. T., & Snedeker, J. (2018). Some inferences still take time: Prosody, predictability, and the speed of scalar implicatures. *Cognitive Psychology*, *102*, 105–126.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press.

Janssens, L., Fabry, I., & Schaeken, W. (2014). "Some" effects of age, task, task content and working memory on scalar implicature processing. *Psychologica Belgica*, *54*(4), 374–388.

Kapa, L. L., & Colombo, J. (2014). Executive function predicts artificial language learning. *Journal of Memory and Language*, *76*, 237–252.

Katsos, N., & Bishop, D. V. M. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*, 67–81.

Kline, R. (1998). *Principles and practice of structural equation modeling*. The Guilford Press.

Kramer, J. H., Mungas, D., Possin, K. L., Rankin, K. P., Boxer, A. L., Rosen, H. J., & Widmeyer, M. (2014). NIH EXAMINER: Conceptualization and development of an executive function battery. *Journal of the International Neuropsychological Society*, *20*(1), 11–19.

Kray, J., & Lindenberger, U. (2000). Adult age differences in task switching. *Psychology of Aging*, *15*(1), 126–147.

Landau, B., & Jackendoff, R. (1993). "What" and "Where" in spatial language and spatial cognition. *Behavioral and Brain Sciences*, *16*(2), 217–238.

Lefcheck, J. S. (2016). piecewiseSEM: Piecewise structural equation modeling in R for ecology, evolution, and systematics. *Methods in Ecology and Evolution*, *7*(5), 573–579.

Lefcheck, J. S. (2019). piecewiseSEM: Piecewise structural equation modeling in R. https://cran.r-project.org/web/packages/piecewiseSEM/vignettes/piecewiseSEM.html

Lewandowsky, S., Duncan, M., & Brown, G. D. A. (2004). Time does not cause forgetting in short term serial recall. *Psychonomicbulletin & Review*, *11*, 771–790.

Loban, W. (1963). *The Language of elementary school children* (Research report no. 1). National Council of Teachers of English.

López-Palma, H. (2007). Plural indefinite descriptions with "unos" and the interpretation of number. *Probus*, *19*(2), 235–266.
MacWhinney, B. (2000). *The CHILDES project*. Lawrence Erlbaum.
Maratsos, M. P. (1974). Preschool children's use of definite and indefinite articles. *Child Development*, *45*, 446–455.
Maratsos, M. P. (1976). *The use of definite and indefinite reference in young children: An experimental study of semantic acquisition*. Cambridge University Press.
Martí, L. (2008). The semantics of plural indefinite noun phrases in Spanish and Portuguese. *Natural Language Semantics*, *16*(1), 1–37.
Marty, P., Chemla, E., & Spector, B. (2013). Interpreting numerals and scalar items under memory load. *Lingua*, *133*, 152–163.
Mayer, M. (1969). *Frog, where are you?*. Dial Press.
McMurray, B., Klein-Packard, J., & Tomblin, J. B. (2019). A real-time mechanism underlying lexical deficits in developmental language disorder: Between-word inhibition. *Cognition*, *191*, 1–13.
Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, 320–334.
Miller, K., Schmitt, C., Chang, H.-H., & Munn, A. (2005, November). *Young children understand some implicatures* [Paper presentation]. Proceedings of the Annual Boston University Conference on Language Development, Boston, MA, United States.
Milsark, G. (1977). Toward an explanation of certain peculiarities of the existential construction in English. *Linguistic Analysis*, *3*, 1–29.
Miyake, A., Friedman, N., Emerson, M., Witzki, A., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100.
Morey, C. C., & Bieler, M. (2013). Visual short-term memory always requires general attention. *Psychonomic Bulletin & Review*, *20*(1), 163–170.
Musolino, J. (1998). *Universal grammar and the acquisition of semantic knowledge* [Doctoral dissertation]. University of Maryland.
Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*(2), 165–188.
Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective and Behavioral Neuroscience*, *5*(3), 263–281.
Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics-pragmatics interface. *Cognition*, *86*(3), 253–282.
Pearl, J. (2009). *Causality* (2nd ed.). Cambridge University Press.
Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation* [Doctoral dissertation. MIT.
Postal, P. (1964). Limitations of phrase structure descriptions. In J. K. a. J. Fodor (Ed.), *Readings in the philosophy of language* (pp. 137–151). Prentice-Hall.
Pouscoulous, N., Noveck, I., Politzer, G., & Bastide, A. (2017). A developmental investigation of processing costs in implicature production. *Language Acquisition*, *14*(4), 347–375.
Pratt, A., Arrieta-Zamudio, A., & Grinstead, J. (2020, November). *Predicting scalar implicature interpretations from lexical knowledge* [Paper presentation]. Boston University Conference on Language Development, Boston, MA, United States.
Pratt, A., Grinstead, J., Leon, N., & Padilla-Reyes, R. (2018). Generation of scalar implicatures in Spanish-speaking adults and children: The role of the Question Under Discussion. *Signos Lingüísticos*, *14*(28), 64–86.
Pratt, A., Grinstead, J., Ortiz-Ramírez, P., Arrieta-Zamudio, A., Cantú-Sánchez, M., Carreto-Guadarrama, X., & Flores-Avalos, B. (2019, November). *Simultaneous vs. Consecutive Action and Implicatures* [Paper presentation]. Boston University Conference on Language Development, Boston, MA, United States.
Rangel, E., Romero, S., & Gómez, M. (1988). *Batería de Evaluación de la Lengua Española para niños de 3 a 11 Años: Manual de Aplicación, Calificación e Interpretación*. Secretaría de Educación Pública,

Dirección General de Educación Especial. [Spanish Language Evaluation Battery for children from 3 to 11 years of age: Manual for application, scoring and interpretation. Public Education Secretariat, Department of Special Education.]
Reinhart, T. (1998, October). *The processing cost of reference set computation: Acquisition of stress shift and focus* [Paper presentation]. UCLA Linguistics Colloquium, Los Angeles, CA, United States.
Reinhart, T. (2004). The processing cost of reference set computation: Acquisition of stress shift and focus. *Language Acquisition*, *12*(2), 109–155.
Ricker, T. J., & Cowan, N. (2010). Visual working memory is disrupted by covert verbal retrieval. *Psychonomicbulletin & Review*, *17*(4), 516–521.
Ricker, T. J., & Cowan, N. (2014). Differences between presentation methods in working memory procedures: A matter of working memory consolidation. *Journal of Experimental Psychology*, *40*(2), 417–428.
Roberts, C. (1995). Domain selection in dynamic semantics. In E. Bach, E. Jelinek, A. Kratzer, & B. H. Partee (Eds.), *Quantification in natural languages* (pp. 661–700). Kluwer.
Roberts, C. (2003). Uniqueness in definite noun phrases. *Linguistics & Philosophy*, *26*(3), 287–350.
Russell, B. (1905). On denoting. *Mind*, *66*, 479–493.
Schoolmaster, D. R., Zirbel, C. R., & Cronin, J. P. (2020). A graphical causal model for resolving species identity effects and biodiversity–ecosystem function correlations. *Ecology*, *101*(8), Article e03070.
Shipley, B. (2000). A new inferential test for path models based on directed acyclic graphs. *Structural Equation Modeling*, *7*(2), 206–218.
Shipley, B. (2002). *Cause and correlation in biology: A user's guide to path analysis, structural equations and causal inference*. Cambridge University Press.
Shipley, B. (2009). Confirmatory path analysis in a generalized multilevel context. *Ecology*, *90*(2), 363–368.
Shipley, B., & Douma, J. C. (2020). Generalized AIC and chi-squared statistics for path models consistent with directed acyclic graphs. *Ecology*, *101*(3), Article e02960.
Shipley, B., & Douma, J. C. (2021). Testing piecewise structural equations models in the presence of latent variables and including correlated errors. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*, 582–589.
Siegler, R. S., & Opfer, J. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, *14*, 237–243.
Smith, C. L. (1980). Quantifiers and question answering in young children. *Journal of Experimental Child Psychology*, *30*, 191–205.
Stalnaker, R. (1974). Pragmatic Presuppositions. In M. Munitz & D. Unger (Eds.), *Semantics and philosophy* (pp. 197–213). New York University Press.
Thorward, J. (2009). *The interaction of contrastive stress and grammatical context in child English speakers' interpretations of existential quantifiers* [Bachelor of Arts senior thesis]. The Ohio State University.
Vargas-Tokuda, M., Gutierrez Rexach, J., & Grinstead, J. (2009). Context and the scalar implicatures of indefinites in child Spanish. In J. Grinstead (Ed.), *Hispanic child languages: Typical and impaired development* (pp. 93–116). John Benjamins.
von Fintel, K. (1994). *Restrictions on quantifier domains* [Doctoral dissertation]. University of Massachusetts.
Wang, S. (2019, September). *Scope and scalar implicatures in child mandarin: The role of working memory* [Paper presentation]. The Generative Approaches to Language Acquisition, Milan, Italy.
Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, *73*(6), 913–934.
Woodman, G. F., Vogel, E. K., & Luck, S. J. (2012). Flexibility in visual working memory: Accurate change detection in the face of irrelevant variations in position. *Visual Cognition*, *20*, 1–28.
Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1–B11.
Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science*, *20*, 423–428.