2017

Children's Quantification with Every Over Time

Athulya Aravind
Massachusetts Institute of Technology

Jill de Villiers
Smith College

Peter A. de Villiers
Smith College, pdevilliers@smith.edu

Christopher J. Lonigan
Florida State University

Beth M. Phillips
Florida State University

See next page for additional authors

Follow this and additional works at: https://scholarworks.smith.edu/phi_facpubs

Part of the Linguistics Commons

Authors

Athulya Aravind, Jill de Villiers, Peter A. de Villiers, Christopher J. Lonigan, Beth M. Phillips, Jeanine Clancy, Susan H. Landry, Paul R. Swank, Michael Assel, Heather B. Taylor, Nancy Eisenberg, Tracy Spinrad, and Carlos Valiente

Children's quantification with *every* over time

Athulya Aravind[1], Jill de Villiers[2], Peter de Villiers[2], Christopher J. Lonigan[3], Beth M. Phillips[3], Jeanine Clancy[3], Susan H. Landry[4], Paul R. Swank[4], Michael Assel[4], Heather B. Taylor[4], Nancy Eisenberg[5], Tracy Spinrad[5] and Carlos Valiente[5]

[1] Department of Linguistics and Philosophy, MIT, Cambridge, MA, US

[2] Department of Psychology, Smith College, Northampton, MA, US

[3] Florida State University, Tallahassee, FL, US

[4] University of Texas Health Science Center at Houston, Houston, TX, US

[5] Arizona State University, Tempe, AZ, US

Corresponding author: Athulya Aravind (aaravind@mit.edu)

This article looks closely at two types of errors children have been shown to make with universal quantification—Exhaustive Pairing (EP) errors and Underexhaustive errors—and asks whether they reflect the same underlying phenomenon. In a large-scale, longitudinal study, 140 children were tested 4 times from ages 4 to 7 on sentences involving the universal quantifier *every*. We find an interesting inverse relationship between EP errors and Underexhaustive errors over development: the point at which children stop making Underexhaustive errors is also when they begin making EP errors. Underexhaustive errors, common at early stages in our study, may be indicative of a non-adult, non-exhaustive semantics for *every*. EP errors, which emerge later, and remain frequent even at age 7, are progressive in nature and were also found with adults in a control study. Following recent developmental work (Drozd and van Loosbroek 2006; Smits 2010), we suggest that these errors do not signal lack of knowledge, but may stem from independent difficulties appropriately restricting the quantifier domain in the presence of a salient, but irrelevant, extra object.

1 Introduction

Since the earliest work on children's behavior with universal quantification, researchers have noticed an odd error they frequently make, called "Quantifier Spreading" or "Exhaustive Pairing". The error can be described as follows. Consider a situation in which there are three cowboys, each riding a horse, and an extra horse without a rider. The sentence in (1) is true in this scenario.

(1)   Every cowboy is riding a horse.

However, given a picture depicting this scenario and asked if every cowboy is riding a horse, children frequently say "No" and point to the extra horse as justification for their negative response. The same sentence in (1) would be false in a scene with four cowboys, only three of them riding horses. Children sometimes make a complementary error where they say "Yes" when asked if (1) is true with respect to such a scene involving an extra, horseless cowboy. This is called an Underexhaustive error, as it involves a failure to find all the relevant cowboys.

This paper presents results from a large-scale, longitudinal study—the first, to our knowledge—investigating how children's exhaustive pairing and Underexhaustive errors relate to each other and to other cognitive factors like memory and executive functioning. We find an interesting inverse relationship between EP errors and Underexhaustive errors over development: the point at which children stop making Underexhaustive errors is also when they begin making EP errors. We suggest that Underexhaustive errors, which are common at early stages in our study, may indicate a stage in development where children have a non-adult, non-exhaustive semantics for *every*. EP errors, which emerge later, and remain frequent even older stages, are argued to result from particular pragmatic difficulties that arise in the presence of an extra object that is contextually salient, but truth-conditionally irrelevant.

2 Background

While children have been observed to make both EP and Underexhaustive errors, the prevalence and robustness of the two error types are unequal: EP errors have been shown to be robust across age—they persist until around age 9 (Roeper 2004)—and across languages—e.g. Dutch (Drozd & van Loosbroek 2006), Turkish (Freeman & Stedmon 1986), Japanese (Sugisaki & Isobe 2001), Catalan (Gavarró & Escobar 2002), French (Inhelder & Piaget 1964), and Russian (Kuznetsova et al. 2007). Underexhaustive errors do not occur with the same frequency or persistence. As a result, much of the literature on children's understanding of universal quantification has focused on explaining EP errors. A number of accounts have been forwarded, but they can be broadly grouped into two main classes of explanation: those that posit a difference in grammatical knowledge between children and adults (the *Partial Competence* accounts) and those that argue that children have full, adult-like understanding of universal quantification, with methodological or processing considerations leading to non-adult performance (the *Full Competence* accounts).

Among the Partial Competence accounts, one view holds that children have an incomplete syntactic representation of quantifier scope. Roeper and de Villiers (1991) hypothesized that in child grammar, the quantifier *every* might "spread" beyond its DP-restrictor, sometimes taking sentential scope. In such cases, the meaning children posit for these sentences might resemble a sentence with a quantificational adverb like *always*. A version of this proposal can be found in Philip (1995). The basic idea here is that children initially misattribute to the universal quantifier *every* an event-quantificational semantics, which we do find in adult grammar for adverbs like *always* or *usually*. Thus, children's representation for the sentence in (1) might be as in (2):

(2)   For every event *e* such that *e* is an event in which either a cowboy or a horse participates, or *e* is a potential subevent of a cowboy-riding horse event *e'*, a cowboy is riding a horse in *e*.

Under this approach, both the *accurate* rejection of (1) in Underexhaustive scenarios (which would depict an extra cowboy) and the *erroneous* rejection of the sentence in over-exhaustive scenarios (with an extra horse) could stem from the same underlying problem: in neither situation is every subevent a cowboy-riding-horse event.

The weak quantification approach of Drozd (2001) is similar in spirit. He suggests that children might interpret *every* in a highly context-dependent manner, as is often the case with certain weak quantifiers. The particular parallel he draws is between children's EP errors and a reading for sentences with the weak quantifier *many* that was first identified in Westerståhl (1984). Consider the sentence in (3) and the two available interpretations for the sentence paraphrased in (3a) and (3b):

(3)   Many Scandinavians are Nobel Prize winners
        a.   Many people from Scandinavia have won the Nobel Prize
        b.   Many winners of the Nobel Prize are Scandinavians

The reading in (3a), in which the first argument of *many* serves as its restrictor, is obviously false—our world knowledge tells us that it is implausible that a lot of Scandinavians have won the Nobel Prize. The preferred reading for the sentence, however, is what is often called the "switched reading" in (3b), which might indeed be true. Many different proposals have been put forth to account for this reading, with some involving literal switching of the restrictor and nuclear scope (e.g. Herburger 1997) and others relying on pragmatic mechanisms to derive the meaning (e.g. Cohen 2001). What is crucial here, however, is that this reading is not possible with strong quantifiers. Drozd suggests that this is precisely where child and adult grammars diverge: children are able to give *every* a weak-quantifier construal, which allows them, in the right contexts, to access a "switched reading" for these sentences.

   Geurts (2003) also invokes the weak-strong distinction among quantifiers in his account. His analysis of children's interpretation of quantified sentences discusses a mapping problem between syntactic structure and semantic representation. Unlike the approach taken in Drozd (2001), the relevant piece is the relational and non-intersective nature of strong quantifiers, contrasted with the non-relational, intersective nature of weak ones. Strong quantifiers, Geurts argues, involve more intricate Logical Forms and are more costly to process than weak quantifiers. In child grammar, the interpretation of *every* in (1) is initially given a strong construal, but the mapping may fail because of the overall complexity involved. At this point, instead of constructing the set of x's such that x is a cowboy in the restrictor position, the child restricts the quantifier via discourse factors alone, as if it were a weak quantifier. The adult interpretation in which both semantic construal and mapping work correctly, is given in (4). The elements inside the front brackets represent the domain of the quantifier *every*.

(4)   [x: x is a cowboy] <every> [y: y is a horse, x rides y]

However the child's semantic interpretation begins as in (5), where the domain contains a variable that may be filled in by the context.

(5)   [… : …] <every> [x,y: x is a cowboy and y is a horse, x rides y]

This means that what is salient in the context could potentially restrict the domain of quantification for children, even when it is syntactically unfeasible. In other words, the child could very well interpret a sentence as in (1) as being about cowboys or about horses. In a context where the horses are salient, the reading might be:

(6)   [y: y is a horse] <every> [x: x is a cowboy, x rides y]

Philip (2012) also suggests that children rely too much on the context, but for him, such context-dependency doesn't entail a weak-cardinal analysis of *every*. Rather, he suggests that children use different methods than adults to restrict the domain of the quantifier. The salience of the symmetry-breaking extra horse, e.g., in the EP scenario leads children to accommodate an unseen cowboy who *should have been* riding this horse.

   All the accounts above in one way or another argue that children do not always identify the appropriate quantificational domain for *every*. These proposals contrast with the Full Competence models, which maintain that children have a fully adult-like representation

of universal quantification, their errors being prompted by extra-grammatical factors. For instance, Crain et al. (1996) argue that children's EP errors are an artifact of infelicitous testing procedures, which failed to satisfy the appropriate usage conditions on Yes-No questions. Specifically, a Yes-No question is felicitous only if there is some possible outcome other than the one represented in the picture (the Condition of Plausible Dissent), a condition that cannot typically be met in the case of out-of-context Yes/No questions. Upon satisfying these felicity requirements, Crain and colleagues report finding a considerable decrease in EP errors.

Other proposals blame the oddity of the visual and/or discourse contexts. Both Freeman et al. (1982) and Brooks and Sekerina (2005) point to the possibility that the near one-to-one correspondence between the two sets could divert children's attention to the unpaired object. Freeman et al. suggest that this could encourage the child to construe the test situation as being about the violation of a one-to-one correspondence. The authors observe that adults, too, sometimes behave like an EP error-making child when there is a naturally expected pairing that isn't met. For example, when shown a picture with cups on saucers and an extra, cup-less saucer, adults were prone to say "No" to the question, "Is every cup on a saucer?" For Brooks and Sekerina, the extra object in the visual context is distracting enough to lead to a cognitive over-load: the salience of the extra object demands attention, exhausting the limited memory and processing resources available to the child. The child then constructs simpler, underspecified representations for the sentence and then relies on contextual clues to solve the task, a strategy they call "Shallow Processing". Thus, Brooks and Sekerina's account is reminiscent of Geurts (2003), but the bulk of the blame is placed on the experimental set-up itself.

Yet others have pointed to the connection between the visual context, in particular the unpaired object, and topicality (Hollebrandse 2004; Drozd & van Loosbroek 2006; Smits 2010). In the absence of explicit contextual support, the extra object, by virtue of being the most visually salient element, is taken to be the discourse topic. Hollebrandse, for instance, tested Dutch children on sentences with *alle* 'every' in both the classic Yes-No question paradigm with and without contextual set-up making clear which set constituted the topic. He found that children who gave EP-answers with uncontextualized *every* sentences did not make the errors when the universally quantified constituent was also the discourse topic. The precise relationship between topicality and quantification domain is not discussed, but a possibility is that the discourse-topic can sometimes help children identify the relevant domain of quantification (e.g. establish that the only cowboys relevant are the ones in the scene or the story). However, this still leaves open the question of why children, but not adults, need further linguistic or contextual cues to narrow down the domain of quantification for *every*.

The experimental evidence amassed over decades of investigation paint a complex picture, but debates about underlying factors aside, it is uncontroversial that children make errors with universal quantification when they encounter certain tasks. The puzzle is compounded by the fact that children make errors with *every* at stages in development where they seem to otherwise show sophisticated knowledge of key properties of the quantifier. For instance, by 5 years of age, children have been shown to know that *every* is downward-entailing in its restrictor (Gualmini et al. 2003), and that it shows the definiteness effect, a characteristic of strong quantifiers (Meroni et al. 2007).[1] To gain a better understanding of the nature of children's underlying knowledge of the quantifier *every* at different points in development, this paper explores the time-course in development

---

[1] The definiteness effect refers to the phenomenon where certain constructions, e.g. the English existential *there*-construction, do not permit strong quantifiers as their theme argument.

of the two main types of errors——EP errors and Underexhaustive errors. Consider again what the different theories might say about the relationship between EP errors and Underexhaustive errors. In Philip (2004; 2012), Underexhaustive responses were not given any linguistic account and were considered a separate cognitive error. This position is maintained in Roeper, Strauss and Pearson (2006) who classed Underexhaustive responders as "perseverators" or "yes-sayers". Geurts' (2003) weak quantifier account with its open restrictor position predicts that EP errors and Underexhaustive errors should co-occur as they are due to the same underlying process. Unlike the other theories, Geurts explains Underexhaustive errors by the same mechanism: if there are cats without apples, the child might say "Yes", because every apple *is* being held by a cat. Notice that this very flexibility makes it impossible to ascertain the precise meaning children assign to *every*, though Geurts himself takes it to be adult-like.

A more direct investigation of the relationship between the two error types were carried out by Altreuter and de Villiers (2006), who looked at both comprehension and production of sentences with *every*. In a first session, 64 children (aged 5–8) answered Yes-No questions about examples of scenes, some designed to maximize EP errors in which characters were lined up so discrepancies stood out, say three cats carrying apples and one cat carrying a banana. The day after testing comprehension, the same children were tested again in production. The subjects were reminded that the day before, some of the computer-narrated sentences had not matched the pictures. The children were then shown a Powerpoint presentation of new pictures similar to the ones used in comprehension and told to make "A true sentence that starts with *every* […]". The study had as its goal to explore whether the errors in comprehension were largely due to processing, or to competence with *every*, by testing whether children made the same errors in production. The results revealed that the children who made Underexhaustive errors in comprehension did significantly worse at the same scenes in production than other types of responders. In other words, their errors in comprehension carried over into production. For these children, the fact that not every cat was holding an apple in the picture was not a problem: it was enough that most of them were, so they freely said, "Every cat is holding an apple". This was not an occasional error but a major form of response for these children. The fact that this error appeared in production suggests that the Underexhaustive errors are neither a consequence of reflex-like "yes-answering" (the cognitive error) nor attributable to weak processing, but perhaps reflective of a non-adult interpretation of the quantifier's meaning. The data on EP errors in Altreuter and de Villiers (2006) were too slight to be of significance, both in comprehension and production. However, it was apparent that there was no correspondence between EP errors and Underexhaustive errors.

The present study builds on this work by examining the developmental profile of both EP and Underexhaustive error types in the same children *over time*. By looking at the phenomenon longitudinally, we hope to better understand (i) when children make and stop making these errors and (ii) how the two error types relate to each other (e.g. are they concurrent?). In this way we might find ways to reconcile some of the disputes above and shed light on the process by which children achieve both the correct construal of the exhaustive meaning and the correct restrictor of *every*.

3 Experiment 1: Children

Our child data are part of a large, longitudinal study on cognitive, linguistic and socio-emotional development. The study was conducted between 2006 and 2011. Each participant was tested 4 times on all the same materials. We were thus able to track the time-course of development for the relevant phenomena. Here we describe only the relevant subset of the materials used in the study overall.

3.1 Participants

The data are from 140 children from subsidized schools in Texas and Florida, a subset of the whole chosen because they completed testing at all four time points. Participants were recruited as part of a study by the School Readiness Research Consortium (Landry 2010, 2014; Lonigan 2015). The majority of the children came from low-income families and were eligible for free school lunches. Participants were 4-years-old at the beginning of the study (mean age, T1 = 4.22) and between 6 and 7 at the end (mean age, T4 = 6.73).

3.2 Materials and design

All of our critical items were transitive sentences involving the universal quantifier *every* in subject position and an indefinite DP with the article *a* in object position. The study included two items involving EP scenarios, in which there was an extra object (EP items henceforth, Figure 1), and two items involving Underexhaustive scenarios, with an extra subject (Underexhaustive items, Figure 2). The small number of critical items was due to time-constraints imposed by the large battery of tests run with each child. We used the classic Yes-No question paradigm, without additional linguistic context. Each child encountered the same scenarios and questions all 4 times she was tested.
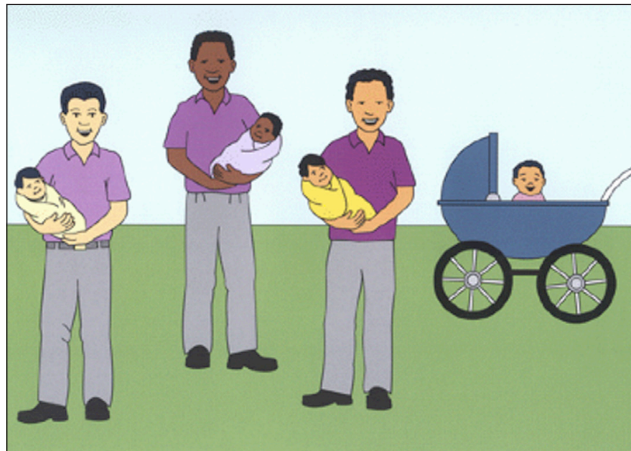
Figure 1: Is every daddy holding a baby?
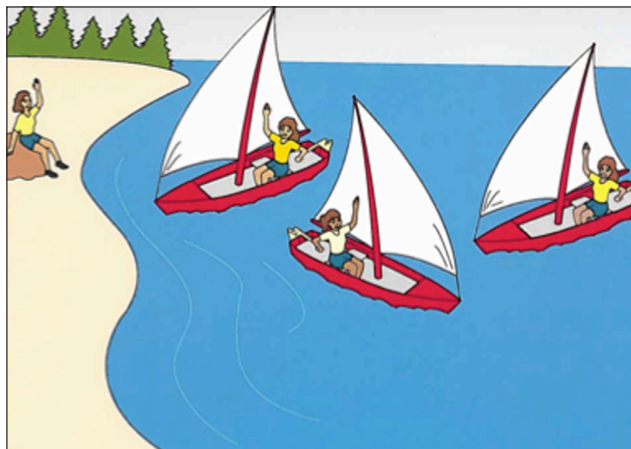
Figure 2: Is every woman sailing a boat?

We also report on a few other linguistic and cognitive measures from the much larger battery that we think are potentially relevant to the quantifier task. Non-verbal IQ (Pattern Analysis sub-test of the Stanford-Binet IQ test) and Verbal Memory (Word Span subtest, CTOPP: Wagner, Torgesen & Rashotte 1999) were measured once for all participants towards the beginning of the study. Additionally, all participants were assessed twice (Times 1 and 2) on inhibitory control ("Knock Tap Test": Korkman, Kirk & Kemp 1998), Vocabulary (Expressive One Word Picture-Vocabulary test: Brownell 2000), and Syntax (*Wh-Questions* sub-test; the DELV Language Assessment: Seymour, Roeper & de Villiers 2005).

3.3 Results

3.3.1 Errors over time

A detailed summary of our results is presented in Table 1. Means and confidence intervals for accuracy across participants at each testing time are given in Figure 3. Recall that there were 2 exhaustive pairing items and 2 Underexhaustive items, so the maximum score a child could score is 2 for each type, and chance is 1.

	EP items				Underexhaustive items			
	Mean	**SD**	**# of children w/ both correct (n = 140)**	**# of children w/ both incorrect (n = 140)**	**Mean**	**SD**	**# of children w/ both correct (n = 140)**	**# of children w/ both incorrect (n = 140)**
T1 (M = 4.22)	1.64	0.63	100	11	0.51	0.75	22	91
T2 (M = 4.75)	1.12	0.86	64	47	1.06	0.92	63	54
T3 (M = 5.8)	0.79	0.85	43	72	1.39	0.82	85	30
T4 (M=6.73)	0.56	0.80	89	27	1.86	0.44	126	5

Table 1: Mean accuracy across children over time.
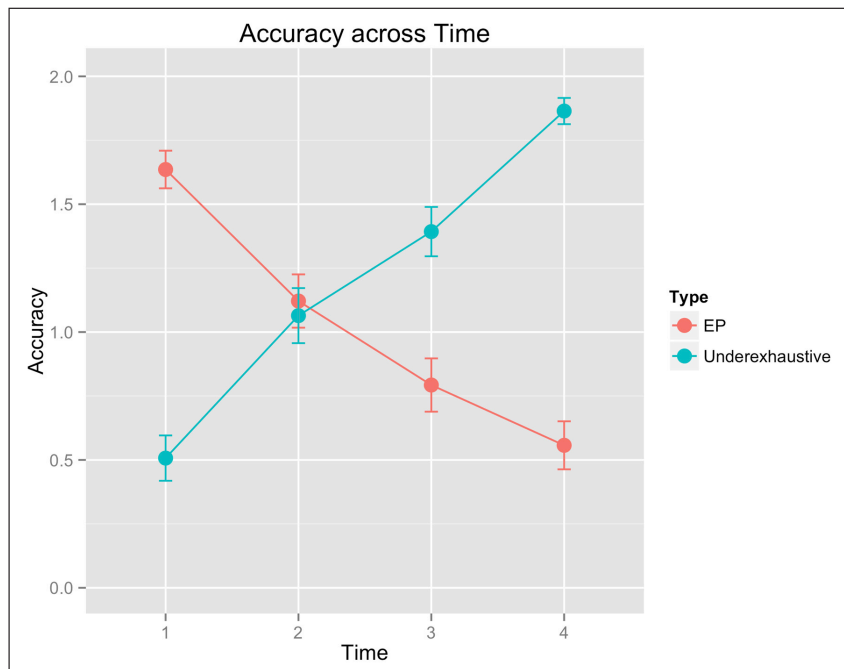
Figure 3: Mean accuracy for the two types across time.

We observe that the developmental trajectories for the two error types look strikingly different. Performance on the Underexhaustive items shows a familiar developmental path: as children mature, they make fewer errors of this kind. Children are at ceiling by Time 4. The trajectory for the EP items is the inverse of this: accuracy appears to *decrease* with time, and at the last stage of testing, children are performing well below chance. This brings us to an important observation: at Time 1, which children are around 4-years-old, most of them (100 out of 140) appear to be getting both EP items correct. At the same stage, many (91 out of 140) are getting both Underexhaustive items incorrect.

To explore these patterns statistically, we conducted a mixed-effects ordinal regression with accuracy as the dependent measure, Type (EP versus Underexhaustive), Time, Nonverbal IQ and Verbal Memory as independent measures; we also included by-item random intercepts and random slopes for the relatedness of Time and Participant. The regression analysis confirmed the trends we observe in the figures above: we find a significant Type * Time interaction ($\beta = -1.97$, $p < .001$). We also found that Type significantly interacted with Nonverbal IQ ($\beta = -0.13$, $p = .005$) and Verbal Memory ($\beta = -0.26$, $p < .001$). In contrast to the Underexhaustive items, the odds of scoring in a higher category for EP items dropped as Time, Nonverbal IQ and Verbal Memory increased.

Notice that the two item-types differed also in the polarity of the correct answer: whereas the adult-like response for Underexhaustive items is "No", with EP items, it is "Yes". Could the observed interaction be explained as the result of a Yes-bias, which dissipates over time? Though we did not have a control condition with *every* that would address this concern, we did have other items involving the Focus Particle *only* which made use of the same Yes-No Question paradigm and required a "No" response to get the answer correct. We reasoned that a child who might be taken to have a Yes-bias would consistently and incorrectly answer "Yes" to these items at the same time as they answer "Yes" incorrectly in the Underexhaustive condition. To ensure that our trends persisted even after taking into consideration the possibility of a Yes-bias, we focused attention on the subset of 58 children who did not consistently say "Yes" on the "only" items.[2] Figure 4 represents accuracy rates on the two conditions across Time for this subset of children. We observe the same general trend, though the patterns are less extreme. A mixed-effects ordinal regression model, parallel to the one fit for the entire sample, demonstrates that the relationships found in the larger sample largely persist. Crucially, we find a significant interaction of Type and Time ($\beta = -1.64$, $p < .001$). We also find an interaction between Type and Verbal Memory ($\beta = -0.36$, $p < .001$). The interaction between Type and Nonverbal IQ was, however, no longer significant in this subsample.

3.3.2 Relationship between the errors

When we look across the board, we find that the trajectories of EP errors and Underexhaustive errors are essentially opposites of each other. The relationship between the two error types was assessed statistically by estimating two additional mixed-effects ordinal regressions. We asked whether performance on one type of items predicts performance on the other. In our first regression model, we included performance on EP items as the dependent measure and performance on Underexhaustive items as a predictor. In the second, performance on EP items was included as a predictor of performance on Underexhaustive items. Both models also included Time, Nonverbal IQ and Verbal Memory as co-predictors, as these factors were found to be significant in our earlier model. For both EP and Underexhaustive items, performance on the other type was a significant *negative* predictor. A unit

---

[2] It is important to note that this is a highly conservative and rigorous measure: it could very well be that children's incorrect "Yes" responses, even when consistent, are due to non-adult linguistic representations and not merely from a bias.
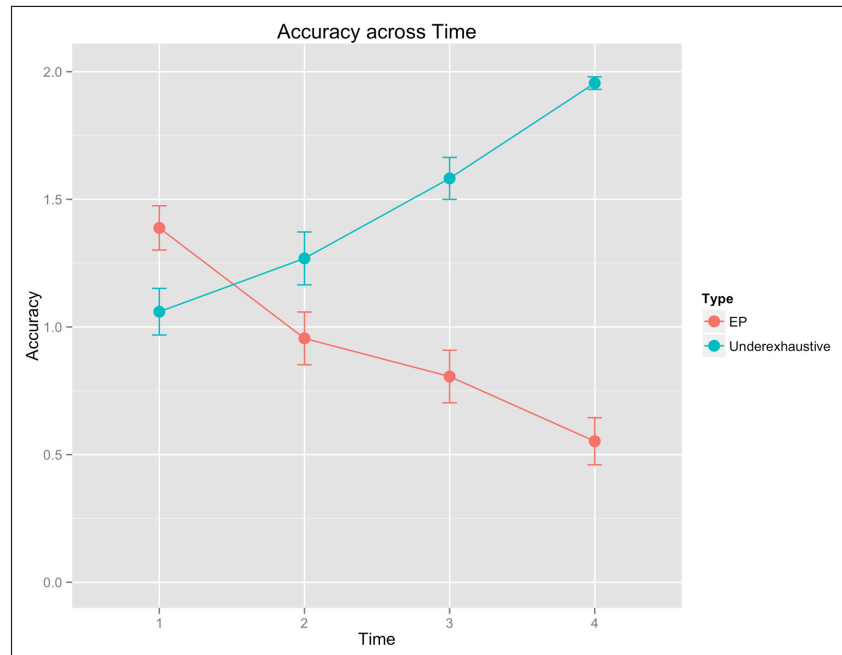
Figure 4: Mean accuracy across time, non-"yes-sayers".

increase in accuracy on Underexhaustive items decreased a child's odds of scoring higher on EP items by 0.12 ($\beta = -2.14, p < .001$). A unit increase in accuracy on EP items decreased the odds of scoring higher on Underexhaustive items by 0.10 ($\beta = -2.32, p < .001$). The striking correlation that emerges is the following: the stage when a child *stops* making the Underexhaustive error is also when she *starts* making EP errors.

Again, to ensure that these trends hold up once we take into account "Yes-biases", we fit the same regression models on the subset of 58 children we had established were not merely "yes-sayers". As with the larger sample, we find that performance on Underexhaustive items is a negative predictor of performance on EP items ($\beta = -1.83, p < .001$), and vice versa ($\beta = -2.09, p < .001$). These results, together with earlier findings by Altreuter and de Villiers (2006) on errors in production, suggest that the inverse relationship between the two error types is genuine and cannot be fully attributed to various extra-linguistic biases.

3.3.3 Relationship to other cognitive factors

We had coincident measures of Vocabulary, Syntactic Ability and Inhibitory Control at Times 1 and 2. To evaluate whether these factors influence performance on either type of quantification questions, we consider cross-sectional data from just Time 2, at which there was the greatest variance in performance. A multiple regression analysis once again reveals significant interactions between item Type and the other factors. As Vocabulary scores increase, the odds of scoring higher on EP items *decrease* by 0.94 ($\beta = -0.07, p = 0.02$). Similarly, as Syntax scores increase, the odds of scoring in a higher category on EP items decrease by 0.66 ($\beta = -0.41, p = .001$). These patterns suggest to us that EP errors are progressive in nature and are likely not driven by a lack of linguistic sophistication. With inhibitory control, however, we find a different trend: children with higher Executive Functioning are *more* likely to score higher on EP items, but this effect just approaches significance ($\beta = 1.6, p = .09$).[3]

---

[3] A reviewer asks why Executive Functioning was merely a trend, despite our large N. The lack of effect could be due to the small number of critical items tested (2 in each condition) and we might expect a stronger effect were we to test a larger battery of items. We leave this for future work as Executive Functioning does not play a major role in our theorizing.

3.4 Discussion

Our study replicates earlier findings in the literature that children aged 4–7 make errors in comprehension when encountering Yes-No questions with a universally quantified subject. The EP error, in which children erroneously reject a universally quantified statement in the presence of an unpaired object in the visual array, developed around age 5 (Time 2 in our study) and seems to become more prevalent as the child gets older. Our findings suggest that children as old as 7 are making these errors. The complementary Underexhaustive error, in which children erroneously accept a universally quantified statement although not every member of the restrictor set satisfied the relevant predicate, occurred frequently at early stages of development, but decreased with time. Very clearly, the two error types do not reflect the same phenomenon.

Children at the earlier stages of testing make errors on the Underexhaustive items, but performance on these items gets better over time. By Time 3 (mean age = 5.8), only 21% of participating children are getting both of the Underexhaustive items wrong. The steady increase in accuracy on this type suggests that between 5 and 6 years of age, children learn something critical about universal quantification, namely that the property denoted by the nuclear scope must hold for every single member of the set denoted by the restrictor. However, this improvement on Underexhaustive items is accompanied by a simultaneous drop in performance on the EP items. By Time 4, 64% of the participating children got both the EP items wrong, a massive jump from just 8% at Time 1. Do they exit one non-adult stage of universal quantification (as indicated by errors on Underexhaustive items at Time 1), only to enter another one? Another possibility is that EP errors are not a reflection of non-adult semantics for the quantifier, but are indeed driven by extra-grammatical factors, as argued by researchers adopting the Full Competence approach. To investigate this possibility, we conduct an adult control study, discussed in detail in the next section.

4 Experiment 2: Adults
4.1 Participants

Sixteen college-aged undergraduate students (all female) from Smith College and Wellesley College were recruited to participate in the study for either course credit or for no compensation. One participant was excluded from the analysis because she failed to meet the inclusion criterion of 50% overall accuracy.

4.2 Materials and design

The materials were similar to those used in the child task. The scenarios all involved sets of objects in near one-to-one correspondence. The classic Yes-No question paradigm, as in the child study, was used, but with an additional time pressure component: participants were told to respond fast, but accurately. A timeout window was set at 6000 milliseconds to avoid extreme RTs. There were 4 items of the EP type, 4 items of the Underexhaustive type and 8 filler items. The sentences were presented on-screen and the participants responded by pressing two keys on the keyboard associated with "Yes" and "No". Accuracy and Response Time information were collected using the OpenSesame experiment presentation software.

4.3 Results

Mean accuracy and response times for both item types are presented in Tables 2 and 3 respectively. Note that we only consider response times for accurate trials.

The first relevant observation is that adults are making a substantial number of errors on the EP items, compatible with those made by adults in Brooks and Sekarina (2005). The error-rate for the EP type is considerably higher than that for the Underexhaustive

Type	Mean	SD
EP	76%	43%
Underexhaustive	93%	25%

Table 2: Accuracy.

Type	Mean	SD
EP	3404	1392
Underexhaustive	2584	1091

Table 3: Response Times (ms.).

items. Adults' relative difficulty on EP items is evident also in their response times: it takes adults longer to accurately respond "Yes" to EP items than to accurately respond "No" to Underexhaustive items. Due to the small sample size and large variance in adult behavior, the difference in accuracy was not statistically significant. However, a mixed-effects linear regression shows a marginal effect of item Type on Log RTs ($\beta = 0.27$, $p = 0.068$).

4.4 *Discussion*

The interesting finding from Experiment 2 is that adults, too, make errors with universally quantified sentences, and that these errors are asymmetric: adults are more likely to erroneously reject a statement with a universally quantified subject when the scenario displays an unpaired extra object than they are to erroneously accept a universally quantified statement with respect to a scene with an extra subject. We would not want to say that the reason for adult errors is lack of semantic knowledge. What, then, is leading them to make mistakes? One possibility is what Brooks and Sekerina (2005) propose as part of their Shallow Processing Hypothesis: errors are due to cognitive overload due to the salience of the distracting extra objects. We concur with these authors that the salience of the extra object plays a role. In the EP items, the outlier item is, deceptively, *irrelevant* to the truth of the sentence, but participants could nevertheless fixate on the extra object and be led astray.[4] The shallow processing account does not, however, explain the asymmetry in reaction times: EP items take longer for adults to evaluate than the Underexhaustive items, though both involve extra elements in the visual array. In the following section, we discuss another possibility, namely that the EP scenarios, but *not* the Underexhaustive scenarios, involve a violation of principles governing cooperative communication, a pragmatic infelicity that incurs additional processing costs in adults.

5 General discussion

We presented findings from a large-scale, longitudinal study on children's quantification with (subject) *every* and a control study with adults. We were interested in examining how children's EP errors and Underexhaustive errors develop over time and also how the two error types relate. We found, as in previous studies, that children frequently make EP errors and these errors persist into the early school years. In contrast, the Underexhaustive errors occurred primarily in the early stages of our testing. There were two main findings that we think are novel. The first concerns an early stage in development where children do not make EP errors, but *do* make Underexhaustive errors. The second is regarding the

---

[4] It is also possible that the relevance of the salient extra object needs to be inhibited exclusively in the EP condition, where it does not play a role in truth-evaluation, and it is this inhibition that is costly. We are grateful to an anonymous reviewer for this suggestion.

developmental trajectories of the two types of items: they show an inverse relationship. In this section, we examine these two findings in turn.

5.1 A non-adult acquisition stage

At Time 1, when children are 4 years old, we find that they are near ceiling on EP items. At the same time, children make errors on the Underexhaustive items, saying "Yes" to universally quantified sentences when the property denoted by the nuclear scope does not hold for all the individuals in the extension of the restrictor. As we saw earlier, it is not sufficient to argue that this stage is the product of biases to say "Yes". This finding is also inconsistent with many of the previous accounts of EP errors. The Full Competence accounts, for example, cannot straightforwardly account for the early apparent accuracy. If it was simply a matter of meeting conditions of Plausible Dissent, for instance, why should children fail to display sensitivity to this requirement before a certain age? The Shallow Processing account also predicts something different. If it was a matter of cognitive overload, we expect that the younger children be *more* distracted by the extra object in the EP condition and make *more* errors, given that younger children's cognitive resources like working memory and attention are likely more limited than those of older children. The Partial Competence accounts fare no better. The Event-Quantification account of Philip (1995), for instance, would wrongly predict accurate performance on Underexhaustive items. Geurt's (2003) proposal would predict a parallel trajectory for the two errors, a hypothesis that is clearly disconfirmed.

An alternative line of explanation, which would be consistent with both children's high accuracy on EP items as well as their errors on Underexhaustive items, is that children at this stage in development genuinely lack an adult-like understanding of universal quantification. In particular, children at this might instead have a weaker, non-exhaustive meaning for *every*. For instance, it is possible that at early stages, children might assign to *every* a plural existential quantifier meaning: they may accept a situation with respect to a universally quantifier statement "every X is Y" as long as there are multiple Xs that are Ys represented in the scene.[5] This view would also be consistent with findings by Heizmann (2012), who shows that exhaustivity is delayed in acquisition across a range of constructions, including wh-questions and cleft constructions.[6]

5.2 Under-informativity and EP errors

The second important finding concerns the developmental trajectories themselves. Children made more Underexhaustive errors at Time 1 and 2, but the prevalence of this error type steadily decreased over time. With EP errors, on the other hand, we find the opposite trend: whereas children initially appeared to be performing well on the EP items, their performance decreases just as they stop making Underexhaustive errors. At Time 4, when children are at ceiling on the Underexhaustive items, they are *below* chance level on the EP items. The progressive nature of these errors is corroborated by the fact that children who are high-performers on other linguistic measures were more likely to be EP

---

[5] A prediction made by this account is that a child in this stage will not differentiate between a prompt involving *every* and a prompt involving a plural DP. We thank an anonymous reviewer for pointing this out to us.

[6] One might wonder how a child who posits an existential semantics for *every* might get out of this stage, since in upward-entailing contexts, the existential scenario will be true every time the universal scenario is true. We suspect that the crucial evidence will come from downward-entailing contexts. If the child were to encounter a sentence like "Not every cowboy is riding a horse" in a situation where there were three horse-riding cowboys and an additional cowboy, she should realize that her semantics for *every* will need to be revised.

error-makers, and that adults also frequently make similar errors in the same sort of task. This strongly suggests that EP errors cannot have a purely linguistic underlying source.

So what is at the heart of the problem? As previously mentioned, a number of researchers have noted that a universal statement, when paired with visual scenes as those used in our EP condition, feels infelicitous (Crain et al. 1996; Hollebrandse 2004; Drozd & van Loosbroek 2006; Smits 2010). One way of characterizing this infelicity is to say that such sentences are under-informative relative to the conversational goals at hand. Let us consider why this might be. A natural assumption given an experiment task like the one used here is that the goal is to evaluate a description of the visual scene presented. Put differently, it is natural to assume that the universal statement is provided in response to a general question of the form: "What is happening in this picture?" Note, however, that a description like "Every father is holding a baby" is *under-informative* relative to a scene like in Figure 1 and a question of the form above. This is because the visual array contains a baby who is not being held and this extra figure is contextually relevant. A statement that makes no mention of this entity only provides a partial description of the scene.

There has been accumulating evidence within language acquisition research that children are not only highly sensitive to violations of conversational principles, but their response to infelicity may be qualitatively different from those of adults (Hamburger & Crain 1982; Crain & Thornton 1998; Gualmini et al. 2008; Hackl et al. 2015). It is plausible that children's rejection of universal statements in EP-contexts is a direct response to the under-informativity of such statements given the context. If this hypothesis is on the right track, then we expect adults, too, to be sensitive to the resulting infelicity. Adults' lower accuracy rates and longer response times in Experiment 2 give preliminary indication that this might in fact be the case. Of course, adults have means of recovering from such infelicities. For instance, upon hearing the test sentence, they might infer that the topic of inquiry is not the visual scene as a whole, but a proper subset (e.g. the fathers in Figure 1). However, this sort of accommodation requires a great deal of pragmatic sophistication, which a primary-school-aged child, whose experience with deliberately uncooperative conversational settings may be limited, may not possess.

Support for such a view comes from a range of more recent developmental work (e.g. Smits 2010; Philip 2012), which manipulated the perceptual salience and contextual relevance of the extra object in the array, eliciting radically different behavior from children. For instance, when the extra object is made perceptually less salient, or when its irrelevance for the topic of discussion is established in the preceding discourse, children make EP errors at much lower rates.

6 Conclusion

Examining children's quantificational errors over time, we found that there is an interesting inverse relationship in development between Exhaustive Pairing (EP) errors and Underexhaustive errors. This rules out the theory that both errors derive from the same basic failure to properly identify the restrictor. We argue further that the Underexhaustive error is not just "yes-saying", nor can it be ruled a cognitive error, but reflects instead the child's initial (mis-)understanding of *every* as a plural existential quantifier. EP errors, however, persist well into the primary school years and arguably even into adulthood, leading us to conclude that they are not indicative of linguistic failure. Rather, we suggested that they may stem from the under-informativity of the test sentences given the visual context. While adults, too, have difficulties with EP-contexts, quantitative differences between adults and children that point to interesting, potentially non-adult ways in which semantics and pragmatics interface in early language.

Abbreviations

EP = Exhaustive Pairing

Acknowledgements

Competing Interests

The authors have no competing interests to declare.

References

Altreuter, Emily & Jill de Villiers. 2006. Does every child produce "every" correctly? In Tanja Vignevjic (ed.), *Current issues in first language acquisition* (University of Massachusetts Occasional Papers)*,* 1–20. Amherst, MA: GLSA.

Brooks, Patricia & Irena Sekerina. 2005. Shortcuts to quantifier interpretation in children and adults. *Language Acquisition* 13(3). 177–206. DOI: https://doi.org/10.1207/s15327817la1303_2

Brownell, Rick (ed.). 2000 *Expressive one-word picture vocabulary test: Manual.* Novato, CA: Academic Therapy Publications.

Crain, Stephen & Rosalind Thornton. 1998. *Investigations into universal grammar.* Cambridge, MA: The MIT Press.

Crain, Stephen, Rosalind Thornton, Carole Boster, Laura Conway, Diane Lillo-Martin & Elaine Woodams. 1996. Quantification without qualification. *Language Acquisition* 5(2). 83–153. DOI: https://doi.org/10.1207/s15327817la0502_2

Drozd, Kenneth. 2001. Children's weak interpretations of universally quantified questions. In Melissa Bowerman & Stephen Levinson (eds.), *Language acquisition and conceptual development,* 340–376. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/cbo9780511620669.014

Drozd, Kenneth & Erik van Loosbroek. 2006. In Veerle van Geenhoven (ed.), *Semantics in acquisition,* 115–140. Dordrecht: Springer.

Freeman, Norman, Chris Sinha & Jacqueline Stedmon. 1982. All the cars – which cars? From word meaning to discourse analysis. In Michael Beveridge (ed.), *Children thinking through language,* 52–74. London: Edward Arnold.

Freeman, Norman & Jacqueline Stedmon. 1986. How children deal with natural language acquisition. In Ida Kurcz, Grace Shugar & Joseph Danks (eds.), *Knowledge and language,* 21–48. Amsterdam: Elsevier North-Holland.

Gavarró, Anna & Linda Escobar. 2002. A pilot study of quantification in child Catalan. Report de Recerca GGT-02–3, UAB.

Geurts, Bart. 2003. Quantifying kids. *Language Acquisition* 11(4). 197–218. DOI: https://doi.org/10.1207/s15327817la1104_1

Gualmini, Andrea, Sarah Hulsey, Valentine Hacquard & Danny Fox. 2008. The question–answer requirement for scope assignment. *Natural Language Semantics* 16(3). 205–237. DOI: https://doi.org/10.1007/s11050-008-9029-z

Hackl, Martin, Ayaka Sugawara & Ken Wexler. 2015. Question-answer (in)-congruence in the acquisition of *only. Proceedings of the 39th BUCLD.* Somerville, MA: Cascadilla.

Hamburger, Henry & Stephen Crain. 1982. Relative acquisition. In Stan Kuczaj (ed.), *Language development, syntax and semantics,* 245–274. Hillsdale, NJ: Lawrence Erlbaum.

Heizmann, Tanja. 2012. *Exhaustivity in questions and clefts; and the quantifier connection: A study in German and English.* Amherst, MA: University of Amherst Massachusetts dissertation.

Hollebrandse, Bart. 2004. Topichood and quantification in L1 Dutch. *International review of applied linguistics in language teaching* 42(2). 203–215. DOI: https://doi.org/10.1515/iral.2004.010

Inhelder, Barbel & Jean Piaget. 1964. *The early growth of logic in the child, classification and seriation.* London: Routledge and Kegan Paul.

Korkman, Marit, Ursula Kirk & Sally Kemp. 1998. *NEPSY: A developmental neuropsychological assessment.* San Antonio, TX: Psychological Corporation.

Kuznetova, Julia, Maria Babyonyshev, Jodi Reich, Lesley Hart & Elena Grigorenko. 2007. The acquisition of quantifiers in Russian. In Alyona Belikova, Luisa Meroni & Mari Umeda (eds.), *Proceedings of the 2nd conference on Generative Approaches to Language Acquisition North America (GALANA)*, 224–232. Somerville, MA: Cascadilla Proceedings Project.

Landry, Susan. 2009. Enhancing early academic, social, and language competencies: Impacts from an integrated Pre-K curriculum. Presented at the *Society for Research in Child Development (SRCD) 2009 annual conference,* Denver, CO.

Landry, Susan, Tricia Zucker, Heather Taylor, Paul Swank, Jeffrey Williams, Michael Assel, April Crawford, Weihua Huang, Jeanine Clancy-Menchetti, Christopher Lonigan, Beth M. Phillips, Nancy Eisenberg, Tracy Spinrad, Jill De Villiers, Peter De Villiers, Marcia Barnes, Prentice Starkey & Alice Klein. 2014. Enhancing early childcare quality and learning for toddlers at risk: The responsive early childhood program. *Developmental Psychology* 50. 526–541. DOI: https://doi.org/10.1037/a0033494

Lonigan, Christopher, Beth Phillips, Jeanine Clancy, Susan Landry, Paul Swank, Michael Assel, Heather Taylor, Alice Klein, Prentice Starkey, Celene Domitrovich, Nancy Eisenberg, Jill de Villiers, Peter de Villiers, Marcia Barnes & The School Readiness Consortium. 2015. Impacts of a comprehensive school readiness curriculum for pre-school children at risk of educational difficulties. *Child Development* 86(6). 1773–1793. DOI: https://doi.org/10.1111/cdev.12460

Meroni, Luisa, Andrea Gualmini & Stephen Crain. 2007. The strength of the universal quantifier in child language. In Alyona Belikova, Luisa Meroni & Mari Umeda (eds.), *Proceedings of the 2nd conference on Generative Approaches to Language Acquisition North America (GALANA)*, 277–284. Somerville, MA: Cascadilla Proceedings Project.

Philip, William. 1995. *Event quantification in the acquisition of universal quantification.* Amherst, MA: University of Amherst Massachusetts dissertation.

Philip, William. 2012. Acquiring knowledge of universal quantification. In Jill de Villiers & Tom Roeper (eds.), *Handbook of generative approaches to language acquisition*, 351–394. New York, NY: Springer.

Roeper, Tom & Jill de Villiers. 1991. The emergence of bound variable structures. In Thomas Maxfield & Bernadette Plunkett (eds.), *Papers in the acquisition of WH* (University of Massachusetts occasional papers), 225–265. Amherst, MA: GLSA.

Roeper, Tom, Uri Strauss & Barbara Pearson. 2006. The acquisition path of quantifiers: Two kinds of spreading. In Tanja Heizmann (ed.), *Papers in language acquisition* (University of Massachusetts occasional papers), 97–128. Amherst, MA: GLSA.

Seymour, Harry, Tom Roeper & Jill de Villiers. 2005. *Diagnostic evaluation of language variation.* San Antonio, TX: The Psychological Corporation.

Sugisaki, Koji & Miwa Isobe. 2001. Quantification without qualification without plausible dissent. In Ji-Yung Kim & Adam Werle (eds.), *The semantics of underrepresented language in the Americas* (University of Massachusetts occasional papers), 97–100. Amherst, MA: GLSA.

Wagner, Richard, Joseph Torgesen & Carol Rashotte. 1999. *Comprehensive test of phonological processing.* Austin, TX: PRO-ED, Inc.

Westerståhl, Dag. 1984. Determiners and context sets. In Johan van Benthem & Alice ter Meulen (eds.), *Generalized quantifiers in natural language,* 45–71. Dordrecht: Foris Publications.