Acquisition of Numerals, the Natural Numbers, and Amount Comparatives

Martin Hackl, Ella Apostoaie, and Leo Rosenstein

1. Introduction

Acquiring the natural numbers (ℕ) is famously (and somewhat surprisingly, given its rather modest formal complexity) a protracted and laborious process. Using children's understanding of number words (henceforth numerals) in the Give-N task as proxy, Wynn (1990, 1992) and much subsequent work has shown that learning proceeds strictly incrementally for the first three or four numerals until around 4ya, which is when children seem finally able to generalize the association between numerals and cardinalities. One tell-tale sign of this newly gained knowledge is that they now understand the "Cardinality Principle", CP, which identifies the cardinality of any set with the last numeral used when counting the elements of that set. That is, they now understand that the linguistic sequence of numerals stands in correspondence to the order among quantities, such that numerals which are later in the sequence reference larger quantities. However, knowledge of the Cardinality Principle, CP-knower status henceforth, does not guarantee complete knowledge of ℕ. Indeed, what aspects of ℕ CP-knowers do and do not (necessarily) know about is still an open question.[1]

The present paper aims to contribute to this literature with an experimental study that probes CP-knowers' command of the additivity of ℕ. We find that CP-knowers do not yet have robust knowledge of this property, even for numbers that are within the range they are highly familiar with. We suggest that this might be so because being a CP-knower does not entail an appreciation of a property of ℕ

*Martin Hackl, Ella Apostoaie, Leo Rosenstein, Massachusetts Institute of Technology, Department of Linguistics & Philosophy. hackl@mit.edu (Hackl). We are grateful to all the children and adult participants, and to parents, teachers and day care supervisors for their cooperation , as well as the Living Laboratory at the Museum of Science, Boston. We would like to thank Athulya Aravind and all of the members of the MIT Language Acquisition Lab for their comments and feedback, and our undergraduate research assistants for their help with data collection, without which this project would not have succeeded. Thanks also to reviewers and attendees at the BUCLD45 conference for their valuable comments and suggestions.

[1] See Sarnecka and Cary (2008), Davidson et al. 2012 a.o. for discussion. See e.g. Spelke (2017) for arguments that counting may not be constitutive to knowledge of ℕ.

 Proceedings of the 45th annual Boston University Conference on Language Development, ed. Danielle Dionne and Lee-Ann Vidal Covas, 276-289. Somerville, MA: Cascadilla Press.

which is canonically characterized via Peano's Induction Axiom.[2] The Induction Axiom ensures that the successor function does not skip any element of $\mathbb{N}$. I.e., it guarantees that every natural number other than 1 is the successor of some other natural number. In turn, this property guarantees the desired uniform quantitative interpretation of the distance between any two adjacent natural numbers, and so affords us the ability to define an algebraic structure over $\mathbb{N}$, $\langle \mathbb{N}, 1, \geq, +, \times \rangle$, which exhibits all the desired properties of the natural numbers.

To investigate at what point children come to understand the additivity of the natural numbers, we examined their understanding of a set of closely related amount comparative constructions, (1). (1a,b) provide baselines for our critical sentence type in (1c), an amount comparative with a numeral differential which explicitly refers to the quantitative distance between two reference cardinalities.[3]

(1) a. The rabbit has more cabbages than the hedgehog.
    b. The rabbit has more than three cabbages..
    c. The rabbit has two more cabbages than the hedgehog.
    d. max(d. The rabbit has d-many cabbages) = max(d. The hedgehog has d-many cabbages) + 2

The truth-conditions expressed by (1c) require that the number of cabbages that the rabbit has is greater than the number of cabbages that the hedgehog has, by a count of 2. They can be rendered as in (1d), which reflects more closely the logical form that sentences like (1c) are canonically assumed to have, (Heim 2006, a.o.). (1d) makes transparent that deriving the correct semantics of sentences like (1c) involves determining not only the two reference quantities of the comparative operator - the maximal degree such that the rabbit has at least that many cabbages and the maximal degree such that the hedgehog has at least that many cabbages - but also the exact quantitative distance between those two quantities, which is integrated via the addition operator.[4] Important for our purpose is that bare numerals that serve as differential phrases, such as *two* in (1a), need to be interpreted as referring to the number 2 for the construction to give rise to the correct truth-conditions.[5] Thus comparatives with bare numerals as differential phrases provide a direct test of whether learners take numerals to refer to specific

---

[2] Induction Axiom: If P is a property of natural numbers such that (a) 1 has property P, and (b) whenever a natural number has property P so does its successor, then all natural numbers have property P. (Dasgupta 2014:30)

[3] Arii et al. (2017) is, to our knowledge, the only other place where sentences like (1c), labeled there *measure phrase comparatives*, have been studied in child language. See Section 4 for a comparison with our study and further discussion.

[4] Employing subtraction to calculate the difference between the two reference cardinalities produces of course an equivalent formulation. Whether it is cognitively more plausible than appealing to addition is unclear to us and requires further study. (1b) has, of course, the advantage that $\mathbb{N}$ is closed under addition but not under subtraction.

[5] One way to ensure this while maintaining that the basic meaning of numerals is lower-bound is to obligatorily exhaustify differential phrases.

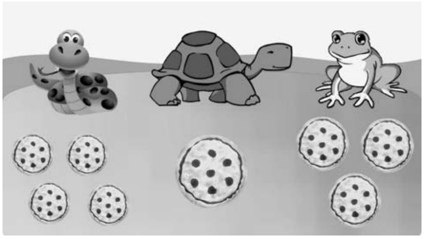
natural numbers and whether they know that the set they belong to, $\mathbb{N}$, is closed under addition.

2. The child experiment
2.1. Methods and materials

Our experiment consisted of 2 parts: first, each participant completed a give-N task in order to determine their "knower-level", the highest number they could reliably identify. Our main experiment took the form of a truth value judgment (TVJ) task staged with a puppet, (see Crain & Thornton 1998). We presented children and a hand puppet with a series of short stories involving three animals and sets of objects of various sizes. The participant would be instructed that the puppet needs help with practicing numbers, and that if he says something right or something silly, they will have to explain to the puppet why what he said was right or silly. Each story ended with a picture in which all of the animals in the story had a different number of items: one, three or four. All of these quantities and corresponding numerals are within the competence range of CP-knowers, as demonstrated with the Give-N task. The items were sized to take up similar amounts of the screen regardless of quantity, and the three and four quantities were always separated by the one, in order to make them more visually distinct, as in Figure 1. The puppet would then utter a series of sentences describing aspects of the scene and the child would be asked, for each sentence, to tell the puppet if they thought that he was right or silly, and why. The experiment consisted of a total of nine stories; one practice, three fillers[6], and five targets. Truth-values were controlled so that within each sentence category there were as many true as false items resulting in an overall ratio of 50 true : 50 false across all items.

As shown in Figure 1, target items consisted of a set of four sentences, conditions A, B, C, and D, presented as descriptions of the same scene. Conditions A and B serve as baselines. Condition A used a basic comparative construction which does not use a numeral to make explicit reference to a relevant quantity. Accurate comprehension indicates knowledge of the syntax and semantics of comparatives as well as the ability to reason with the relevant quantities, specifically appreciating that the domain of quantities forms a weak order and so supports ordering statements involving the ≥-relation. Condition B employed a comparative quantifier where the standard of comparison is given by a numeral, (Hackl 2000). Accurate comprehension indicates understanding that numerals reference quantities and that the canonical sequential order among numerals corresponds directly to the ≥-relation among the associated quantities.

[6] The fillers were similar to the target stories, except that each animal had a different type of object, distributed similarly to the distribution patterns of target items. Fillers only had two sentences with truth values pseudo-randomized across the experiment and adopting a simple structure such as "The spider has sunglasses." These broke up the monotony of similar story endings and allowed us to differentiate between children with non-adultlike judgments and children who were not paying attention to the task.

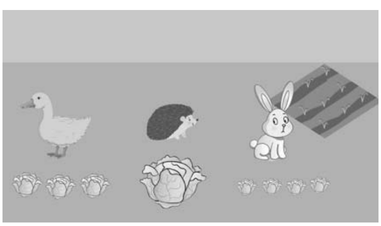


A. The snake has more pizzas than the frog. (True)
B. The tortoise has more than two pizzas. (False)
C. The snake has one more pizza than the frog. (True)
D. The frog has two more pizzas than the tortoise. (True)

Figure 1. Sample target item, Type 1

Conditions C and D combine the critical elements of A and B into the same sentence using the format of a comparative with a differential. In this construction the numeral does not provide the standard of comparison, which remains implicit just like in the basic comparative, but refers to the difference between the two reference quantities that are compared. Accurate comprehension therefore relies not only on commanding the $\geq$-relation among quantities and the semantic import of the order among numerals, it crucially also requires being able to determine the quantitative distance between the two reference quantities. Thus accurate comprehension rides on representing quantities in an algebraic structure that supports additivity, (see Krantz et al. 1971). Applied to the domain of natural numbers, it implies command of the "+" function, which, in turn implies knowledge of the induction axiom.

We included two versions of the differential comparative in order to distinguish between adultlike and possible non-adultlike parsing strategies. For instance, a parsing strategy that ignores the differential would lead participants to ignore the numeral entirely in favor of a simple comparative interpretation, which would make their responses indistinguishable from adultlike judgments on true C and D sentences; a snake who has one more pizza than the frog naturally has more pizza in general, so a "right" response based only on "more" would inflate accuracy measurements. In the same situation the snake does not have two more, so the same parsing strategy would reveal the underlying reasoning as non-adultlike. C sentences are set up so that "one more" would be true and "two more" would be false with respect to the picture, while D sentences are set up in the reverse way, as in Figure 2.



B. The hedgehog has more than two pizzas. (False)
A. The rabbit has more cabbages than the duck. (True)
C. The rabbit has two more cabbages than the duck. (False)
D. The duck has one more cabbage than the hedgehog. (False)

Figure 2. Sample target item, Type 2

The order in which conditions were presented with each story was pseudorandomized.[7] Each sentence type was associated with an equal number of scenarios that made it true and false, pseudo-randomly distributed throughout the experiment, to avoid biasing participants into perseverating on either yes or no answers. Moreover, four different orders of our items were employed to avoid any bias based on what truth values come first in the experiment.

Each experimental session started with a Give-N task to gauge participants' competence with the numbers employed in our task, as well as to determine our population of CP-knowers. We asked participants to put a certain number of star toys from one basket into another, starting with 1. When the child was done moving stars, the experimenter would confirm with, "Is that one?" When a child got a number correct, the experimenter would ask for the next number up. If the child got a number wrong, the experimenter would ask for the next one down, and so on until either the participant got a number wrong twice (and the number below it correct at least twice), in which case we considered them to be a "knower" of the number below the one they got wrong twice, or the participant was correct twice on five, at which point we considered the child a CP-knower.

2.2. Participants

Our participant pool consists of a total of 68 English-speaking children ages 3-6 (range: 3;0 - 6;5; mean: 4;4) recruited from and tested at daycares, preschools and Museums in the Boston area. Of these, eight were excluded for perseveration of Yes or No answers, and a further two were excluded for getting 50% or more of the filler sentences wrong. Of the remaining 58, two were 2-knowers (one 3ya, one 4ya), six were 3-knowers (four 3ya, one 4ya, one 5ya), four were 4-knowers (two 3ya, two 4ya), and 46 were CP-knowers (seven 3ya, 23 4ya, 14 5ya, 2 6ya). Our analyses are focused on the CP-knower population.

2.3. Results

Figure 3 displays the mean accuracy rates of CP-knowers for target sentences broken down by truth-value. We observe high accuracy rates for A and B, as well as for the false variants of C and D, but not for their true counterparts.

We examine this pattern statistically to assess the effect of sentence type and truth-value using a linear mixed-effects logit model, maximally specified on the condition of convergence. We find no main effect of sentence type, but a significant interaction ($p < 0.001$) between sentence type and truth due to the contrasting accuracy rates in both C and D conditions. We furthermore examined whether accuracy rates for our different sentence types in a given truth-value is different from chance (using again a linear mixed effects model maximally specified within converging parameters).

---

[7] A and B sentences were always presented before C and D sentences, but A was not always before B and C was not always before D.

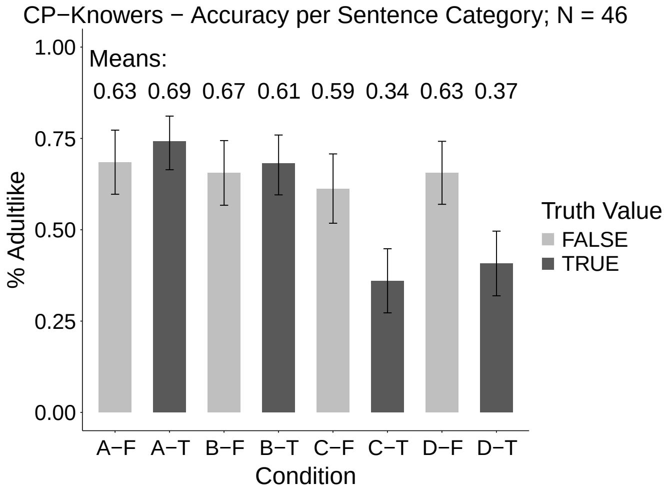


Figure 3. Mean accuracy rates for CP-knowers across sentence types and truth-values

We find that every condition's mean accuracy was significantly different ($p < 0.05$ for all) from chance, except for C-T whose significance was marginal ($p = 0.07$). All conditions were above chance, except for C-T and D-T conditions, which were significantly below chance. This indicates that our participants are not guessing. They provide adult-like responses, except that in C and D cases, they seem compelled to tell the puppet he was being silly regardless of the truth value.

2.4. Discussion

Two aspects of our results are worth highlighting: 1. Accuracy rates on basic comparatives and comparative quantifiers are high for both adult-true and adult-false variants, suggesting adult-like understanding of these sentence types. 2. the effect of truth on accuracy rates for comparatives with differentials is unusual. One normally finds a "yes"-bias in TVJ tasks, i.e., children tend to be charitable and say "true" in case of uncertainty rather than "false". However, our C and D data indicate the opposite bias at work. A "false" bias for these sentences accounts for both higher, above-chance rates of adult-correct responses in false conditions, and lower, below-chance rates in true conditions. In other words, above-chance performance on adult-false versions of C and D does not necessarily indicate adult-like comprehension of these sentences, only systematic "silly" responses.

To understand why children might have a "false"-bias for C/D sentences, we analyzed the justifications our participants offered. Many children, when prompted with true differential comparatives, gave explanations that seemed to indicate that they were confident in their counting abilities, but did not address the differential comparative itself, instead focusing on the counts of items in front of the individuals. For instance in C-true situations (as in Fig. 1) when faced with, "The snake has one more pizza than the frog," participants would respond e.g., "No, the snake has four." or would count out the four pizzas. In a D-true situation (also Fig. 1), where animal A had 3 and animal B had one, and the sentence was "A has two more than B", they would often say, "No, A has 3 and B has 1" or simply "No, he has three." Even more tellingly, in "one more" sentences (C-true and D-false), children would often point to the middle animal, who always had one, and explain that the sentence was silly "because the tortoise is the one who has one pizza." In the C-true cases (such as in Fig. 1), we found 10 cases (out of 148 possible) where the middle animal was brought up as justification for a false judgment, which is striking because C-true sentences do not mention the middle animal at all. In D-false sentences (e.g. in Figure 2), where the animal with 3 was said to have "one more than" the animal who had one, children explained their "silly" answers 21 times (out of 116) by simply pointing out the animal who had one, and another 19 times by explaining that "no, he has three."[8] Most explanations on C and D sentences were just lists of numbers that one or more of the animals had, but sometimes included contrastive pitch-accent on the numeral.[9]

A simple way to understand these justifications is to assume that they are offered as corrective statements, countering not the actual target sentence that was uttered by the experimenter, but the child's interpretation of it, which must have a suitable form and meaning to license it. Specifically, we can understand these justifications if our participants simplified target sentences like (2a) to variants like (2b).[10] (3a), with contrastive pitch-accent on *four,* is completely appropriate as a corrective of (2b). We can similarly understand justifications pointing to the animal in the middle, which does have the quantity of objects referred to by the differential, (3b). Here the contrast is on the subject rather than the numeral.

(2) a. The snake has one more pizza than the frog.
  b. The snake has one pizza.
(3) a. No. (The snake has) FOUR!
  b. No. (It is) the tortoise (who has one)!

Children may have employed other simplification strategies that are less easily detected by their justifications than the one in (2). For instance, children may have, on occasion, ignored the differential phrase in which case a judgment

---

[8] The totals are different because, in cases where participants declined to answer or answered unintelligibly, data points would be excluded without excluding the participant. We also had five target stories per participant, meaning that each participant saw two of one version of each condition (true or false) and 3 of another. This was to be balanced across participants, but data collection was cut short due to the advent of COVID-19.

[9] Each test session was audio-recorded, and the recordings were later reviewed to assess children's adult-likeness by their justifications (as opposed to answers that just happened to match the condition's truth value), as well as to detect pitch-accents.

[10] This conjecture parallels a proposal in Arii et al. (2017) to which we will return in more detail in section 4.

of "true" would appear adult-like for adult-true items and non-adult-like for adult-false items. This may be the source of some of the "misses" for the adult-false variants of C and D and of some of the "hits" for the adult-true ones.

In sum, a closer look at the justifications provides compelling evidence for the claim that the accuracy rates for adult-false C and D items are inflated and do not, in fact, reflect adult-like comprehension of these conditions. Instead, they predominantly seem to be the result of a response strategy based on simplifying the target sentence. Combining this assessment with the observation that their performance on adult-true variants of the C and D categories was below chance leads us to conclude that they do not exhibit adult-like comprehension of comparatives with differentials.

The picture that emerges from our findings seems at first blush quite surprising. We see robust adult-like comprehension of basic comparatives as well as comparative quantifiers and for the very same scene decidedly non- adult-like comprehension of comparatives with differentials. Their adult-like performance on A and B suggests that they have adult-like command of the syntax and semantics of basic comparatives and of measure phrase comparatives where the standard of comparison is provided by a bare numeral. It also suggests that they have the required number knowledge. I.e. they know what it means to have more cabbages than someone else, they know what it means to have one, two, three, or four cabbages, and they know what it means to have more than three cabbages. Nevertheless, they exhibit non-adult-like performance on C and D, which superficially simply combine the elements of linguistic and number knowledge they have just shown to command. Why? What is not simple about combining these elements into the format of comparative with differentials?

Three possibilities come to mind. 1) It could be that comparatives with differentials are simply syntactically too complex for children to parse or process correctly. We partly address this question in our adult control experiment (Section 3) and discuss it further in Section 4.1. 2) Children might lack a critical piece of semantic machinery to compute meanings for comparatives with differentials as Arii et al. (2017) propose. We will discuss this in Section 4.2. 3) They may not have the required number knowledge to interpret comparatives with differentials numerals correctly. This is what we argue for in section 4.3.

3. The adult control

To assess whether the unexpected effect of truth on accuracy rates for C and D sentences is specific to children and thus should have a developmental explanation, we ran an adult control study via Amazon Mechanical Turk. Since we expect adults to have full command of numerals and comparatives with and without differentials as well as the natural numbers, we expect, prima facie, uniformly high accuracy rates across all conditions. However, online testing is known to produce noisier data than in-person testing due to reduced experimental control of the test environment and a general prevalence of satisficing. Thus, a more nuanced expectation would be to allow for lower accuracy rates in conditions that are on general grounds more susceptible to inattention and satisficing errors. For our materials, this applies to conditions C and D since the target sentences are longer and potentially syntactically more complex than the A and B conditions.[11] Moreover, they may require more complex mental computations (calculating the precise difference) than the A and B conditions when assessing their truth-value against a given scene. Thus we may see lower accuracy rates for our C and D conditions as well as lower accuracy rates for false conditions (charity) but, if our child experiment results are indicative of a developmental phenomenon, we should not see a truth-value effect targeting true items and only in our C and D conditions.

To conduct the adult control study we simply transported the materials from our child experiment to IBexFarm, asking for a True/False judgment and then a justification for each sentence for each story, which was written in text below each picture. We ran 89 participants total, of which 32 had to be excluded for lack of attention to the task.[12]

Mean accuracy rates for the remaining 57 participants are shown in Figure 4. Accuracy rates are high across the board and indeed almost identical across conditions, with a minor drop for D-false.[13] Clearly, this is quite different from the pattern we saw in our child experiment. It is also worth pointing out that we don't even see traces of inattention errors or satisficing errors in our adult data. This means that the task was quite easy for our adults, irrespective of condition. Even our longest and (possibly) syntactically most complex items in the C and D conditions which required the most complex mental computation during verification did not yield lower accuracy rates.[14] We take this to mean that the explanation of our child data has to be child-specific.

---

[11] Whether comparatives with differentials are truly structurally more complex than basic comparatives and comparative quantifiers is not immediately obvious and depends on the analysis of these constructions. See e.g. Hellan (1981) for a proposal where basic comparatives have an existentially quantified but silent differential phrase and Hackl (2000) for an analysis of comparative quantifiers as clausal comparatives.

[12] Participants were excluded for getting more than 3/8 fillers incorrect (just like kids), timing out more than 3 times, or giving justifications which indicated spam or lack of English fluency. We believe the high attrition rate is due to the easy short length of the experiment as lower time investments attract low-quality responses.

[13] Statistical examination of these rates with a maximally specified linear mixed effects model reveals an effect of sentence type ($p < 0.05$) attributed to lower accuracy for sentence type D. We argue, however, that this is a spurious effect that has no theoretical significance since the effect disappears ($p > 0.1$) when we pool data from C and D conditions, which, recall, exemplify the same category and are distinguished solely for the purpose of identifying non-adult parsing strategies. Pooling the C and D data from our child experiment does not remove the effect we observed, of course.

[14] We also recorded reaction times and again found no effect of complexity. By all indications available to us our C and D conditions were no more complex or taxing for our adult participants than the A and B conditions.

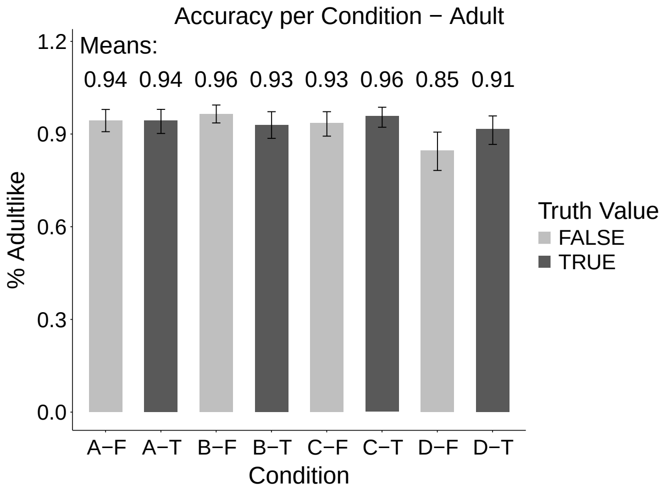


Figure 4. Mean accuracy rates of adults' judgments on target sentences

4. Discussion
4.1. Syntactic complexity

Recall that the inspection of the justifications given by our child participants for target sentences in the C and D conditions lead to the conclusion that they predominantly followed an answer strategy based on simplifying versions of the original sentences like *The snake has one more pizza than the frog* to *The snake has one pizza.* A seemingly simple hypothesis to account for this shortening strategy is to propose that the target sentences were simply too complex for our participants, maybe even just too long. To us this seems quite implausible as the underlying cause for our pattern. To start with, recall that whatever complexity difference there may be between our A and B conditions versus our C/D conditions,[15] it must be so minor that it is not detectable even in extremely noisy test situations such as Amazon's Mechanical Turk platform. Children, nevertheless, would have to be acutely sensitive to this difference - even 5 year olds, since many of our CP-knowers were of that age. Moreover, this proposal would have to identify what exactly it is that makes our sentences too long or too complex. After all, children at this age can produce and comprehend sentences of greater length and comparable or even greater complexity. Without such a precisification this proposal has no explanatory value and with it, whatever it may be, the appeal of its simplicity disappears unless independent reasons can be given for why that particular structural aspect presents such difficulty.

---

[15] See footnote 11.

More damning yet is the fact that children do seem to produce comparatives with differential phrases, quite early in development. Preliminary results from an ongoing CHILDES corpus study reveal examples like the following[16]:

(4) *CHI: “I need some more milk, a little bit more.”
(Gleason, recording Victor, line 686; age 2;05.03)
(5) *CHI: “And he just skips because he knew a lot more than me.”
(Peterson, McCabe, recording 34, line 139; age 6;03)
(6) *CHI: “in a few more weeks (.) winter will be here (.) Mom .”
(Kuczaj, recording 040814, line 158; age 4;08)

The fact that we find cases like (4-6) in child production undermines the syntactic complexity hypothesis. Furthermore, a striking feature of these examples is that in most cases the differential phrases are expressed by vague quantity expressions like *a little bit, a lot, a few,* which are syntactically more complex than bare numerals. Clearly this is unexpected if this type of structure were too complex for children in our age range. In fact, we came across a similar justification in our own experiment, offered by a 3ya 2-knower (who is not included in our analysis):

(7) P: The snake has more pizzas than the frog. (A-True)
CHI: [True] It has more, but a little more. But I know it has more.

Given that children seem to have no real trouble producing comparatives with differentials, especially with differentials that are realized by vague quantifiers like *a little, a lot, a little bit,* we deem a syntactic complexity account for our comprehension data exceedingly implausible.

4.2. A semantic account (Arii et al. 2017)

Arii et al. (2017) presents comprehension data for comparatives with differentials that are rather parallel to ours. Their test environment does not provide base-line data for basic comparatives, (1a), and comparative quantifiers, (1b), however, which makes the data set they get for each participant less comprehensive than ours.[17] Nevertheless, we believe that our observations can be

---

[16] We also found a small number of cases like (i). The productive pattern, however, seems to be generated by vague differentials as in (4-6).

(i) *CHI: and there's only a a few more [= picture in book].
*CHI: let me see how much more there are.
*CHI: there's only two more.
(Gelman, 2014-IndDiff corpus, rec. 39C-R2, lines 286-288; age 4;11.13)

[17] In other respects it is more comprehensive. E.g. in addition to amount comparatives with differentials they also tested comparatives based on *tall* and *long.* Moreover, they ran parallel tests in English and Japanese thereby covering a significant amount of morpho-syntactic variation in how languages express comparative concepts.

seen as a replication of their Experiment 1 - Part 2 results. Arii et al. (2017) diagnoses the poor comprehension in a manner that is quite similar to our conjecture. Specifically, they propose that children cannot assign a meaning to sentences of the form *A has n more Xs than B* and instead predominantly deploy a comprehension strategy of shortening those sentences to *A has n Xs.*

The underlying cause hypothesized in Arii et al. (2017) for why children employ this comprehension strategy is semantic in nature. Concretely, measure phrases, including differential phrases, are assumed to be licensed by a functional head *Meas* which locates the degree denoted by the measure phrase on a measurement scale. In the context of a comparative construction the scale is reset with the standard of comparison denoted by the *than*-constituent serving as the new bottom element of the scale, (Svenonius & Kennedy 2006). The child grammar is hypothesized to have a different specification for *Meas,* however. $Meas_{Child}$ is said to presuppose that the scale has always 0 as its bottom element. Thus the scale cannot be reset to a non-zero bottom element as is required when *Meas* is part of a comparative resulting in the construction being uninterpretable.

We do not have the space to discuss the merits and problems of this proposal as we see them in any detail here. Instead, we simply point out that the proposal is designed to be insensitive to the specific properties of differential phrases. It applies equally to bare numerals as well as vague quantity expressions. Thus, it cannot account for the fact that children produce comparatives with differentials in an adult-like manner when the differential is vague.

Our own 3ya 2-knower in (7) is a case in point. Clearly, CHI is applying perfectly sound, adult-like reasoning about the relevant quantities. At the same time they seem to be admitting that they do *not* know the precise amount that the snake's number is above the frog's. This is entirely unexpected under the Arii et al. (2017) proposal. Indeed, assuming that cases like (7) as well as the corpus data in (4-6) are indicative of a systematic contrast between vague quantity and bare numeral differentials we can extract a strict limit on the space of possible semantic accounts of our and Arii et al.'s comprehension data: the non-adult-like semantics needs to be specific to bare numerals. It cannot extend to vague differentials.[18] The most direct way of developing such an account is to identify the non-adult-like semantic component with a non-adult-like understanding of numerals. In the next subsection we suggest that incomplete knowledge of ℕ, the domain that the interpretation function maps numerals onto, is ultimately responsible for this.

4.3. Additivity of the natural numbers

We have seen that even CP-knowers struggle with interpreting comparatives with numeral differentials in an adult-like manner in our experimental environment, resorting instead to an answer strategy in which the numeral talks about the quantity of items that the animal has, rather than the difference between

[18] We are currently working on a follow-up experiment that compares vague differentials ("a few more" and "a lot more") with numeral differentials ("two more" and "four more") against the same visual context.

the two relevant quantities. This is striking because the quantities we used in our study as well as their respective numeral labels are well within the range that CP-knowers are highly familiar with. Moreover, taking data like (8) at face value, children at this and even younger ages are fully capable of reasoning with differences, just not, it seems, when the difference is denoted with the precision afforded by the natural numbers. This suggests that the difficulty's underlying cause lies in an incomplete knowledge of ℕ rather than being linguistic in nature.

Using the formal characterization of ℕ via the Peano axioms, we ask how we might identify the missing piece of cognitive machinery. One option is that CP-knowers do not yet have access to the Successor Function, as Davidson et al. (2012) seem to suggest (contra Sarnecka & Carey (2008)). We think this is implausible, because CP-knowers do have access to recursive functions, at least within their linguistic system. Moreover, they seem to have learned that the association between numerals and quantities follows a productive rule and that neither domain has a build-in upper bound.[19]

We would like to suggest instead that CP-knowers may not know that the domain of quantities that numerals are mapped onto is governed by a principle that ensures a uniform quantitative difference between any two adjacent members of the domain. That is, they may not know the Induction Axiom. This may be why CP-knowers struggle with bare numeral differentials but not with vague quantity differentials. Comparatives with differentials depend on being able to calculate the difference between quantities. Whether and how such calculations can be performed depends on the algebraic structure of the domain to which the differential phrases are mapped. For vague quantity expressions that structure is plausibly the Approximate Number System (ANS), which affords comparisons as well as arithmetic operations such as addition, subtraction, etc., albeit on quantity representations that do not have sharp boundaries (Dehaene 1997 a.o.).

For numerals, we suggest, the algebraic structure at this stage is an order-preserving transformation on the ANS that "chunks" (quantizes) the ANS into a set of discrete, non-overlapping, ordered intervals. Such a representational system would support a seemingly adult-like interpretation of comparative quantifiers as in sentences like *The rabbit has more than three cabbages* because computing its truth-conditions requires only knowing that the rabbit's quantity of cabbages is ranked above the quantity assigned to *three.* More generally, any form of reasoning with numerals that relies on the ordering of their denotations would be preserved, including contrast inferences, ordering-based entailments, and quantity implicatures (see e.g. Barner & Bachrach 2010, Feinman et al. 2019). Crucially, however, this representational system would not be guaranteed to support additivity inferences because the chunking would not guarantee that no sections of the ANS are skipped or that chunks are of equal size. To ensure this, the learner has to adopt an additional principle governing the output of the chunking transformation, a principle that is well-characterized by the Induction Axiom.

---

[19] CP-knowers' having access to the Successor Function does not guarantee that they can compute its **extension** error-free beyond the range of familiar numbers, allowing for the error behavior that was observed in Davidson et al. (2012).

References

Arii, Tomoe, Kristen Syrett, and Takuya Goro. (2017) “Investigating the form-meaning mapping in the acquisition of English and Japanese measure phrase comparatives”, *Natural Language Semantics* 25.1: 53-90.

Barner, David, and Asaf Bachrach. (2010) “Inference and exact numerical representation in early language development”, *Cognitive Psychology* 60.1: 40-62.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. (2015) “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software*, **67**(1), 1–48. 10.18637/jss.v067.i01.

Crain, Stephen, and Rosalind Thornton. (1998) Investigations in Universal Grammar: A guide to experiments on the acquisition of syntax and semantics. The MIT Press.

Davidson, Kathryn, Kortney Eng, and David Barner. (2012) “Does learning to count involve a semantic induction?”, *Cognition* 123.1: 162-173.

Dehaene, Stanislas. (1997). *The Number Sense: How the Mind Creates Mathematics*. New York, NY: Oxford University Press.

Drummond, Alex. (2012). *Ibex: A Web Interface for Psycholinguistic Experiments 6*. Available online at: https://github.com/addrummond/ibex (Accessed 04/22/2020).

Feinman, Roman, Joshua Hartshorne, and David Barner. (2019) “Contrast and Entailment: Abstract Logical Relations Constrain How 2- And 3-year-old Children Interpret Unknown Numbers”, *Cognition* 183:192-207.

Gelman, Rochel, and C. R. Gallistel. (1978) “The Child's Understanding of Number”, Cambridge: Harvard University Press.

Hackl, Martin. (2000) “Comparative Quantifiers”, PhD diss., MIT.

Heim, Irene. (2006) Remarks on comparative clauses as generalized quantifiers. Ms., MIT. https://semanticsarchive.net/Archive/mJiMDBlN/.

Hellan, Lars. (1981) Towards an Integrated Analysis of Comparatives. Gunter Narr Verlag, Tübingen.

Krantz, David H., R. Duncan Luce, Patrick Suppes, Tversky. (1971). The Foundations of Measurement Vol. 1. Additive and Polynomial Representations. San Diego, London: Academic Press Inc.

Peano, Giuseppe. (1889) The principles of arithmetic, presented by a new method. In *From Frege to Gödel* [78], pages 83–97. English translation of “Arithmetices principia, nova method exposita”, Turin, 1889.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org/

Sarnecka, Barbara W., and Susan Carey. (2008) "How counting represents number: What children must learn and when they learn it.“ *Cognition* 108.3: 662-674.

Spelke, Elizabeth S. (2017) “Core Knowledge, Language, and Number”, *Language Learning and Development*, 13:2: 147-170.

Svenonius, Peter, and Christopher Kennedy. 2006. Northern Norwegian degree questions and the syntax of measurement. In Phases of Interpretation (Studies in Generative Grammar 91), ed. Mara Frascarelli, 131–161. Berlin: Mouton de Gruyter.

Wickham, Hadley. (2016) *ggplot2: elegant graphics for data analysis*. Springer New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Wynn, Karen. (1990) “Children’s understanding of counting”, *Cognition* 36.2: 155-193.

Wynn, Karen. (1992) “Children’s acquisition of the number words and the counting system”, *Cognitive Psychology* 24.2: 220-251.

Proceedings of the 45th annual Boston University Conference on Language Development

edited by Danielle Dionne
and Lee-Ann Vidal Covas

Cascadilla Press Somerville, MA 2021

Copyright information

ISSN 1080-692X
ISBN 978-1-57473-067-8 (2 volume set, paperback)

Ordering information

To order a copy of the proceedings or to place a standing order, contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, sales@cascadilla.com, www.cascadilla.com