# Quantifying kids

## Bart Geurts

Humboldt University, Berlin   &

University of Nijmegen, Department of Philosophy
P.O. Box 9103, 6500 HD Nijmegen, The Netherlands

bart.geurts@phil.kun.nl
www.phil.kun.nl/tfl/bart

## 1. Introduction

It has been known for several decades that young children have difficulties, of various kinds, with universal sentences. In this paper I present an analysis of the main errors that have been reported in the literature. My proposal is based on an old idea, viz. that children's errors are caused by a non-canonical mapping from syntactic form to semantic representation. Previous accounts starting out from this assumption were not entirely successful because they lacked the proper framework for dealing with quantification and domain restriction. Fortunately, everything falls into place, or so I shall argue, if we adopt the treatment of quantification I have developed and defended, on independent grounds, in earlier publications.

## 2. The facts

Stories about language acquisition tend to begin in Geneva, and this one is no exception. Inhelder and Piaget (1959) presented children with displays of coloured squares and circles, and asked their juvenile subjects to assess quantified sentences against these displays. What they found was that the task is a lot harder than one would suspect. Here are a few examples:

(1)  *Scene:* 14 blue circles, 2 blue squares, 3 red squares.
      *Q:* Are all the circles blue?
      *A:* No, there are two blue squares.

(2)  *Scene:* blue circles, blue squares, red squares [no exact numbers given].
      *Q:* And are all these squares red?
      *A:* No, because there is a blue one.
      *Q:* And are all the blue ones circles?
      *A:* Yes.

(3)  *Scene:* 5 blue circles, 3 red squares.
     *Q:* Well, look, are all the circles here blue?
     *A:* Yes … no.
     *Q:* Why?
     *A:* There are red ones.
     *Q:* Where?
     *A:* There are red squares and blue circles.

Similar results have been obtained in countless subsequent experiments, which demonstrated beyond any reasonable doubt that young children often deviate from their elders when it comes to evaluating quantified sentences. Although most experimental work has been done with English-speaking subjects, comparable data have been reported for Dutch (Philip and Verrips 1994, Philip and Coopmans 1995, Drozd and van Loosbroek 1999), Turkish (Freeman and Stedmon 1986), Japanese (Takahashi 1991), Catalan (Philip 1995), Chinese (Chien and Wexler 1989), and of course French (Inhelder and Piaget 1959).

With the help of an artificial example, the nonadult responses described in the literature may be classified as follows. When asked if the sentence 'Every X is a Y' is true or false, a child may claim that the sentence is:

(A)  false in a situation like this:  $\|X\| = \{1, 2, 3\}$    $\|Y\| = \{1, 2, 3, 4\}$
(B)  true in a situation like this:  $\|X\| = \{1, 2, 3, 4\}$  $\|Y\| = \{1, 2, 3\}$
(C)  false in a situation like this:  $\|X\| = \{1, 2, 3\}$    $\|Y\| = \{1, 2, 3\}$    $\|Z\| = \{4\}$

According to this classification (1) counts as a Type-A response, which puts it in the same class with (4) and (5):

(4)  *Scene:* 5 apples and 3 pigs eating 1 apple each.
     *Q:* Every pig is eating an apple … Does this picture go with the story?
     *A:* No. Those two apples have no pig.        (Philip and Takahashi 1991)

(5)  *Scene:* 4 garages and 3 cars, each occupying 1 garage.
     *Q:* All the cars are in the garages.
     *A:* No.                     (Donaldson and Lloyd 1974)

Inhelder and Piaget's example (2) betrays a Type-B error, as does the following:

(6)  *Scene:* 5 cars and 4 garages, each occupied by 1 of the cars.
     *Q:* All the cars are in the garages.
     *A:* Yes.                (Donaldson and Lloyd 1974)

Type-C errors are harder to find (in the literature, that is), but example (3) is a case in point, as is the following:

(7)   *Scene:* 3 cats holding a balloon, and 1 mouse holding an umbrella.
        *Q:* Is every cat holding a balloon?
        *A:* No. (pointing to the mouse)                    (Philip and Verrips 1994)


Problems with quantified sentences have been observed in 3-year olds and persist at least up to age 7. Not all children have them, but many of them do, although it is hard to say how many, because reported error rates vary greatly; this variation makes it pointless to give precise figures, but error rates in excess of 50% are quite common. It is unlikely that all types of error are equally frequent, but unfortunately the literature doesn't yield a clear distribution pattern, because experimenters have tended to confine their attention to Type-A and Type-B errors (and the former category seems to have been especially popular). Indeed, Type-C responses have been used to screen out subjects in pretests, so there is little in the way of systematic data on this category. Nonetheless, such evidence as is available suggests that Type-C errors are much rarer and less persistent than others.

   Children's problems with quantified sentences are sensitive to a number of factors. One important factor is the type of determiner. To begin with, errors seem to be restricted to sentences with universal quantifiers: Takahashi (1991) reports that sentences with cardinal determiners don't cause any trouble, and the same goes for sentences with definite subjects (Drozd 2001), unless they contain 'floated' universal quantifiers (Donaldson and Lloyd 1974). Smith (1979, 1980) found a clear contrast between *some* and *all,* sentences with *some* nearly always prompting adult-like responses, and according to Braine and Rumain (1983), the only problem with *some* is that it occasionally has numerical connotations, leading young children to reject 'Some As are Bs' on the ground that the number of As is 'too large', but such findings are irrelevant in the present context. Hence, the problems described above are confined to sentences with *all*, *every*, and *each*. There may be differences amongst these determiners, as well, but as on this score the empirical record is neither very substantial nor entirely consistent, they will be ignored in the following.[1]

   Not only is there a clear distinction between tasks with universal and existential sentences: problems arising from the former may be compounded by previous exposure to the latter. Smith (1979, 1980) presented 4- to 7-year-olds with quantified questions like 'Are all animals cats?', which had to be resolved not against a visual display but against basic world knowledge (of a kind available already to the youngest subjects). Half of the children started with the batch

---

[1] Drawing on unpublished research by himself, Krämer, and Loosbroek, Drozd (2001) reports that Dutch five-year-olds are sensitive to the distinction between *alle* 'all' and *iedere* 'every', while four-year-olds are not. According to Drozd and Philip (1993) and Philip (1995), on the other hand, English-speaking subjects were indifferent to the distinction between *all* and *every*. On yet another hand, Freeman and Stedmon (1986) found differences between *all*, *absolutely all*, *every*, and *every single*, which only emerge with sufficiently large situations: with a 4-garage/3-car array the choice of determiner doesn't have an effect, but with a 4-garage/5-car array it does. Clearly, this is an issue on which further experimental research is needed.

of *all*-questions, while the other half was first taken through the *some*-questions. Smith's main results were that the first group performed quite well on all tasks, while the second group had considerable difficulties with *all*-questions (questions with *some* proved to be unproblematic). It appears, therefore, that initial exposure to a series of *some*-questions may cause errors with subsequent *all*-questions, but not vice versa.

Another parameter affecting children's responses is the way a situation is laid out and presented. Compare, for instance, Donaldson and Lloyd's examples (5) and (6). It seems as if in these tasks the cars and garages are not on an equal footing: being more natural landmark objects, the garages appear to function as the background against which the cars are examined. This initial impression is reinforced by Freeman et al.'s (1982) finding that a group's relative salience may bias children's responses one way or the other, a finding that was confirmed in experiments by Drozd and van Loosbroek (1999):

(8) *Scene:* 4 cowsheds and 3 cows, occupying 1 cowshed each; cowsheds have been made salient in the preceding discourse.
*Q:* Are all the cows in the cowsheds?
*A:* No.                                                    (Freeman et al. 1982)

(9) *Scene:* as in (8), but now the cows are salient.
*Q:* Are all the cows in the cowsheds?
*A:* Yes.                                                   (Freeman et al. 1982)

In the same vein, Freeman and Stedmon (1986) discovered that group size may be a factor, too: further exploring the car/garage paradigm first introduced by Donaldson and Lloyd, Freeman and Stedmon found that 'the more cars, the more correct answers' (p. 41). A similar improvement may be achieved by keeping the collection not of garages but of cars constant across tasks (Freeman et al. 1982). Taken together, these data suggest rather strongly that, as compared to adults, children fix the domain of quantification with more regard to pragmatic cues and correspondingly less regard to grammatical constraints. If a collection of individuals is particularly salient (for whatever reason), children tend to assume that a given quantifier ranges over it, despite the fact that grammatical constraints bar such an interpretation. This is not yet an explanation, though, as it leaves open a number of questions. For example, it remains to be seen what it means for a grammatical constraints on quantification to be less rigid in children than it in adults, or why this only holds for universal quantifiers.

## 3. *Full grammatical competence?*

The research surveyed in the foregoing has been criticized by Crain and his associates for being based on a flawed experimental design (Crain et al. 1996, Crain and Thornton 1998). Crain et al. allege that earlier studies failed to observe certain pragmatic felicity conditions, which led some children to respond in a way that is out of step with their linguistic competence. In

particular, it is crucial, according to Crain et al., that in a truth value judgment task both answers be should be made available as genuine options:

> In the contexts for yes/no questions, felicitous usage dictates that both the assertion and the negation of a target sentence should be under consideration. (p. 302)

As applied to example (4), say, this principle calls for a scenario in which the possibility that some pigs don't eat an apple is considered, too. If this possibility is left out of account altogether (as it usually would be in experiments prior to Crain et al.'s), children will be puzzled as to why the experimenter wants to know if the statement is true. Crain et al. report that in experiments satisfying this condition children performed about as well as adults, and they conclude that 'young children have full grammatical competence with universal quantification.' (Crain et al. 1996: 83)

Crain and his co-workers are arch nativists, who 'take it to be the *null hypothesis* that children have full linguistic competence' (Crain et al. 1996: 147; emphasis added). Someone taking a more moderate view on language acquisition (or dare we call it language learning?) will find their argument uncompelling, and will probably draw a more guarded conclusion, viz. that Crain et al. have hit upon yet another pragmatic influence on children's response patterns. But even this is contestable (cf. also Drozd 2001, Geurts 2000). To begin with, though it is beyond dispute that there are all sorts of felicity conditions which constrain linguistic processing, it remains to be seen that the one identified by Crain et al. is among them. Contrary to what Crain et al. contend, it is doubtful that a yes/no question is pragmatically infelicitous unless both the affirmative and the negative answer are 'under consideration' in any substantial sense. In my own experience, children are rather good at answering all manner of questions that would be infelicitous according to Crain et al., and there is plenty of experimental evidence to prove the point. For example, as we saw in the foregoing, children have no difficulties assessing universally quantified sentences against basic world knowledge, nor are they troubled by sentences with definite or indefinite subjects. Such data not only indicate that Crain et al.'s felicity condition is a chimera, they also remain unexplained on the theory they present.

A further reason for doubting Crain et al.'s diagnosis is that the parameter they claim to have manipulated in their experiments is open to alternative construals. To see this, consider how (4) would have to be presented in order to be pragmatically felicitous, according Crain et al. This would call for a story which had one or more pigs consider whether or not they should have an apple before all pigs finally decided to eat one. Such a story would inevitably raise the salience of the three pigs, of course, and we knew already that salience may affect children's responses. Hence, it isn't even clear that Crain et al.'s main finding is a new one.

Finally, even if it were true that Crain et al.'s experiments prove children to have 'full grammatical competence' with respect to universal quantification, their diagnosis still leaves open some of the most intriguing issues raised by the empirical facts. Why is it, for example, that older children and adults aren't bothered by an experimental set-up that, by Crain et al.'s lights, is hopelessly flawed? Crain et al. have very little to say about this:

We believe the reason is that older children and adults are better at taking tests than young children. To be successful in previous studies, participants were required to accommodate the fact that the negation of the test sentences was not under consideration in the adult interpretation. Presumably, older children and adults have learned to see through misleading circumstances in which test sentences are presented […]. (Crain et al. 1996: 117)

Needless to say, this is not very helpful, and it doesn't even begin to explain the data reviewed in § 2. Or, to put the same point in a positive way: Crain et al.'s nativist views on budding linguistic competence leave room for a broad range of analyses as to how young children *process* universal sentences, including some proposals that have been made in the literature as well as the proposal to be advanced below.

*4. Quantification and discourse*

The analysis that I have to offer will be implemented within the general framework of discourse representation theory (DRT; Kamp 1981, Kamp and Reyle 1993), and employ the presuppositional account of domain restriction developed by Geurts and van der Sandt (1997, 1999) and Geurts (1999); this section is an ultra-synoptic run-down of the key ideas.

As its name already indicates, DRT is concerned with the interpretation of discourses, not just sentences. As a discourse unfolds, interlocutors incrementally construct a model of what has been said. These models are called 'discourse representation structures' (DRSs). A DRS consists of two parts: a universe of so-called 'reference markers', which represent the objects under discussion, and a set of DRS-conditions which encode the information accumulating on these objects. The following DRS represents the information that there are two individuals, one of whom is a girl, the other a boy, and that the former kissed the latter:

(10)  [x, y: girl(x), boy(y), x kissed y]

The universe of this DRS contains two reference markers, x and y, and its condition set is {girl(x), boy(y), x kissed y}. A DRS like this can be given a straightforward model-theoretic interpretation by defining what it means for a function to be a verifying embedding of a DRS into a given model. A function $f$ embeds (10) into a model M iff the domain of $f$ is {x, y}, $f(x)$ is a girl in M, $f(y)$ is a boy, and $f(x)$ kissed $f(y)$.

Meanwhile it will have become clear that (10) is to reflect the intuitive meaning of:

(11)  A girl kissed a boy …

and indeed, the idea is that, in the absence of any information about the context in which this sentence is uttered, the semantic representation of (11) is (10). So the indefinite subject and object NPs in this sentence cause the introduction of two new reference markers, x and y, and contribute the information that x is a girl and y a boy, and the verb contributes the information

that x kissed y. If a discourse begins with an utterance of (11), the DRS in (10) is constructed, and this DRS acts as the background against which the next utterance is interpreted. Suppose now that (11) is followed by a token of (12a):

(12)  a.  … and then he kissed her back.
      b.  [v, w: v kissed w]

(12b) is the DRS that reflects the semantic content of (12a) before the pronouns are resolved. In this DRS, the anaphoric pronouns *he* and *her* in (12a) are represented by the reference markers v and w, respectively. These reference markers are underlined to indicate that they require an antecedent. Since (12a) is uttered in the context of (10), the next step in the interpretation of this sentence is to merge the DRS in (12b) with that in (10). The result of this merging operation is (13a):

(13)  a.  [x, y, v, w: girl(x), boy(y), x kissed y, v kissed w]
      b.  [x, y, v, w: v = y, w = x, girl(x), boy(y), x kissed y, v kissed w]
      c.  [x, y: girl(x), boy(y), x kissed y, y kissed x]

Given that (12a) is immediately preceded by (11), the subject pronoun *he* is probably intended to be coreferential with *a boy*, while *her* should link up to *a girl*. At DRS level, this is represented by equating the reference markers v and w with y and x, respectively. These operations yield (13b), which is equivalent to (13c). Either DRS is embeddable in a model M iff M features a girl and a boy who kissed each other.

  Quantified determiners are represented in one of two ways, depending on whether they have a weak interpretation or a strong one.[2] Here is an example with a weak determiner:

(14)  a.  Fred fed two lamas.
      b.  [x: Fred(x),  two [y: lama(y), x fed y]]

The intended interpretation of (14b) is that there are two individuals y such that y is a lama and x fed y. More generally, a function $f$ embeds  Q $[u_1, …, u_m: \quad_1, …, \quad_n]$ into a model M iff there are Q-many individuals *a* for which there is a function $g$ that extends $f$, $g(u_1) = a$, and $g$ embeds $\quad_1, …, \quad_m$ into M. Note that, by convention, Q binds the first reference marker in the embedded DRS. That is, Q binds $u_1$; the remaining $u_2, … , u_m$ have existential force by

---

[2] As has become standard in the literature, I assume that the universal determiners and *most* are always strong (which is why, in English, they may never occur in existential *there*-sentences, for example), whilst most other determiners are weak by default, though they may have strong readings, too. Following general usage, I will apply the weak/strong terminology not only to readings but to determiners, as well. Thus, *all* and *most* are strong, while *some*, *a few*, *at least two*, *at most two*, etc. are weak.

default.[3] If a determiner has a strong reading, a non-relational representation such as in (14b) will not do, and we need something more complex:

(15)  a.  Fred fed all lamas.
       b.  [x: Fred(x), [y: lama(y)] all [: x fed y]]

A function *f* embeds    all    into a model M iff all individuals *a* for which there is a function *g* that extends *f*, maps the first reference marker in    onto *a*, and embeds    into M, there is function *h* that extends *g* and embeds    into M. Note that, by convention, Q binds the first reference marker in the DRS on its left.

   In this version of DRT, anaphora is a special case of presupposition, in the sense that presupposition want to be bound just as anaphors want to be bound. Thus, if we replaced *he* in (12a) with *the boy*, it wouldn't make any difference whatsoever. Anaphora is special in that, a few odd cases aside, anaphoric pronouns *must* be bound, whereas in general presuppositions may be construed by way of accommodation, in which case the presupposed content is inserted in the highest accessible DRS, provided the resulting interpretation is in keeping with general pragmatic constraints of plausibility, felicity, and so on. Suppose the hearer doesn't know yet that Fred is married. Then, nevertheless, the speaker may felicitously utter (16a):

(16)  a.  Fred fed all of his wife's lamas.
       b.  [x: Fred(x), [y, z: lama(y), z is x's wife, z owns y] all [: x fed y]]
       c.  [x, z: Fred(x), z is x's wife, [y: lama(y), z owns y] all [: x fed y]]

By hypothesis, the presupposition triggered by *his wife* cannot be bound, hence must be accommodated, and since accommodation in the highest accessible DRS is the preferred option, it is predicted that (16a) will be construed as (16c), and will therefore be understood, *ceteris paribus*, as implying that Fred has a wife, which is of course correct.

   One of the ways presuppositions are triggered is through focusing. For example, 'Fréd kissed Wilma' presupposes that someone kissed Wilma, whereas with (narrow) focus on *Wilma* instead of *Fred*, the same sentence presupposes that Fred kissed someone. I assume, accordingly, that backgrounded material is presupposed; this assumption is essential to my account of examples like (17a):

(17)  a.  Most people visit Berlin [in the spring]$_F$.
       b.  [: [x: person(x)] most [t: x visits Berlin during t, spring(t)]]
       c.  [t: x visits Berlin during t, [x: person(x)] most [: spring(t)]]

---

[3] This encoding of variable binding deviates from the standard DRT convention, which I have adopted on previous occasions. I introduce a different method here because I will be claiming that the grammatical connection between a determiner and its domain of quantification is less rigid in children than it is in adults, and this idea is more difficult to implement with the standard treatment of variable binding.

        d.   [: [x, t: person(x), x visits Berlin during t] most [: spring(t)]]

Here focusing within the nuclear scope of *most* triggers the underlined presupposition in (17b). Supposing that this presupposition cannot be bound, it must be accommodated, but it cannot be accommodated in the main DRS, as shown in (17c), because there the first occurrence of x would not be properly bound (and even if it were, the resulting reading would be rather implausible), and therefore it must be accommodated one level down, resulting in (17d), which says that most people who visit Berlin do so in the spring. This, I take it, is the most natural reading for (17a) to have.

    There is another way in which quantification and presupposition interact: strong quantifiers presuppose their domains. Handling this type of presupposition demands a rather elaborate apparatus that I don't want to go into here, but as these presuppositions will play a role in the following, a brief discussion is in order. Consider example (16a) again. Here the determiner *all* triggers the presupposition that Fred's wife owns lamas. If this information is not given already, it must be accommodated, which might be problematic in this particular case, because hearers may not be prepared to accommodate information that is surprising or potentially controversial, and in our Western culture owing lamas is something out of the ordinary. On the other hand, if it *is* given that Fred's wife owns lamas, the intended interpretation of (16a) will be greatly facilitated, evidently. Speaking more generally, there is a two-way interaction between context on the one hand and quantifiers and their domains, on the other: if a quantifier is construed as ranging over a domain D, the speaker is understood as conveying that he takes D as given; and vice versa: if some collection D is given, and especially if it is currently salient, it may bias the hearer towards construing a quantifier as ranging over D.


*5.  Weak construals of strong determiners*

To return to the main theme of this paper, my proposal is as follows. Children who give non-adult responses to quantified sentences construe the strong determiner as if it were weak: the problem lies in the mapping between syntactic form and semantic representation; it is a parsing problem. More accurately, it starts out as a parsing problem, which is patched by means of pragmatic reasoning of the kind discussed in the previous section.

    I assume that the distinction between weak and strong determiners is not just a linguistic curiosity, but is directly relevant to the ways quantified expressions are processed. This assumption was already implicit in the analysis of quantification outlined above, according to which strong determiners must be represented by relational structures, while weak determiners give rise to non-relational representations by default; a weak determiner demands a relational representation only if it triggers a domain presupposition, as in the following example:

(18)   Some lamas resented being fed by Fred.

If this is understood as referring only to the lamas owned by Fred's wife (say), *some lamas* gets a strong construal, and therefore has a relational representation. However, this is a marked case; as a rule, *some* will be weak.[4]

It follows from this treatment of weak and strong construals of quantified expressions that the former are simpler than the latter. Strong construals are harder not only because their semantic representations are more intricate, but also because they require a more roundabout mapping from form to meaning (cf. Figure 1). When processing a sentence of the form DXY, X and Y must be processed separately if D is strong. But if D is weak, X and Y may be treated as one unit.
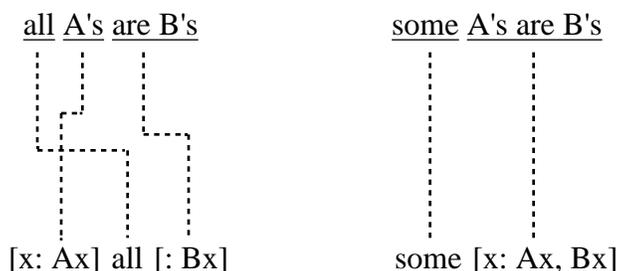
all A's are B's       some A's are B's

[x: Ax] all [: Bx]     some [x: Ax, Bx]

*Figure 1: Mapping surface form to*
*semantic representation:* all *vs.* some

From a semantical point of view, the crucial factor is that weak determiners are intersective and strong determiners are not: in order to ascertain whether DXY is true, if D is weak, we only need to inspect the set $\|X\| \cap \|Y\|$. Hence, the set of lawyers that are crooks is all we need for checking if 'Some lawyers are crooks' is true; any crooks or lawyers outside $\|lawyer\| \cap \|crook\|$ need not be taken into account. Of course, this does not hold anymore if we replace *some* with a strong determiner such as *all*, for strong determiners are not intersective. Therefore, it is not a coincidence that existential sentences afford simpler representations than universal ones: at root, the difference is a matter of content.

The proposed differentiation between existential and universal determiners is supported not only by linguistic data but by experimental findings, too. It has been known for some time that existential sentences are easier to process than universal ones. For example, in an experiment by Just (1974), sentences of the form 'some XY' had shorter response latencies than sentences of the form 'all XY', except when $\|X\|$ and $\|Y\|$ were identical or disjoint, in which cases universal and existential sentences were processed equally fast; the distribution of errors followed the same pattern, i.e. subjects made more mistakes with universal sentences. The same regularities had been observed previously by Meyer (1970) in a rather different experimental set-up, which goes to show that they are fairly robust.

The foregoing considerations show that weak determiners can be handled with simpler means than strong ones, and that it is plausible to assume that they are. But I cannot claim to have proved that it is so, because it is at least possible to generalize to the worst case, and

---

[4] (18) is marked in at least two respects: it has an indefinite subject (subjects are almost always definite) and requires an accent in an usual position, i.e. on *some*.

uniformly treat weak and strong determiners as giving rise to relational interpretations. That is to say, in DRT terms, it is possible to map all quantified sentences onto semantic representations of the form    Q   , regardless whether Q is strong or weak. However, this strategy will not work unless we emend our system of interpretation in other ways as well. In particular, we will have to drop the assumption that a relational quantifier presupposes its domain by definition, and stipulate instead that domain presuppositions are triggered on strong construals only. Thus, strong construals turn out to be more complex than weak ones, after all, although the difference doesn't register in terms of grammar and/or parsing. As far as I can see, there is no knock-down evidence against this line of analysis. However, it seems to me that an account such as mine is more plausible, precisely because it assumes that in general simpler representations and simpler processing strategies will be preferred. Generalizing to the worst case allows for a model that economizes on the *diversity* in representations and processing strategies, which is a way of saving, too, though an arguably less urgent one from a psychological point of view.

   A further argument in favour of the analysis outlined above is that it makes possible a neat account of why children have problems with universal quantification; it is to this account that we now turn. The analysis I propose starts out from the assumption that the child's grasp of the grammar of universal quantification is not deficient in any way. Nor do I see any compelling reason for doubting that children master the logic of universal quantification. In my view, the child's problem (if there is one) lies in the mapping from grammatical form to semantic representation, which is more complicated for universal determiners than it is for most others. Children have a certain tendency to interpret all quantifiers as if they were weak, because it is easier to do so. Thus, if the weak processing strategy is applied to (19a), it is mapped onto (19b) (here and in the following I leave off the DRS's outer brackets):

(19)  a.  Every boy is riding an elephant. (Drozd and van Loosbroek 1999, Drozd 2001)
      b.  every [x, y: boy(x), elephant(y), x rides y]

(19b) hardly counts as a full-blown construal of (19a), of course, because *every* requires two arguments, and I assume that the child knows this; that is to say, the determiner's lexical meaning is transparent enough, it is just the mapping from form to meaning that goes awry. Hence, the child's target representation is something like (20):[5]

(20)  [ … : … ] every [x, y: boy(x), elephant(y), x rides y]

(20) is the child's semantic representation before pragmatic processing sets in. This representation leaves the domain of quantification underdetermined (as compared to the adult's construal), so that there is more elbow-room for pragmatic inferencing.

---

[5] The transition from (19b) to (20a) may look a bit *ad hoc* but it need not be because an analogous transformation is needed for weak determiners to get strong construals.

According to the presuppositional treatment of domain restriction outlined above, the domain of (20) will be constrained depending on the focus/background division within the nuclear scope, which in its turn is at least partly determined by the current context. If the child considers that a given set of boys is currently the most salient discourse entity, he will assume that [x: boy(x)] is backgrounded, and will therefore interpret (19a) as (21a). This interpretation is correct by adult standards, though it has been obtained by non-canonical means. If, on the other hand, the child's attention is focused on a given set of elephants, [y: elephant(y)] is backgrounded, and the same sentence is interpreted as (21b):

(21)  a.  [x: boy(x)] every [y: elephant(y), x rides y]
       b.  [y: elephant(y)] every [x: boy(x), x rides y]

The reading in (21b) will prompt a Type-A response if every boy rides an elephant and some elephants are *sans* boy; it will prompt a Type-B response if every elephant is ridden by a boy and some boys are *sans* elephant; and it will produce the correct response if it so happens that all elephants are ridden by a boy and all boys are riding an elephant.

It will be clear, I trust, how the current proposal accounts for the main facts as discussed in § 2. First, it explains why children's errors are restricted to sentences with strong determiners: this is because strong determiners require a more difficult mapping from form to meaning. Secondly, it explains why errors with universal sentences may be caused by previous exposure to existential sentences: this will happen if the processing strategy for weak determiners, which is easier anyway, is primed by a series of existential sentences, and carried over to subsequent universal sentences. Thirdly, it explains how and why pragmatic reasoning contributes to children's construals of universal sentences: pragmatic reasoning can play a larger role in children than in adults whenever an incorrect mapping from surface form to semantic representation leaves the quantifier's domain underdetermined, but the pragmatic mechanisms that take over at this point are the same in children as in adults.

Thus far I haven't said anything about Type-C errors. How are they accounted for? This is a hard question, though not for lack of possible answers; it is just that the dearth of data doesn't really allow us to distinguish good answers from bad ones. For this reason, the following remarks are quite tentative. One way of looking at Type-C responses is that they are no different from Type-A and Type-B responses. To explain how, let us go back to example (7), which I repeat here for convenience:

(22)  *Scene:* 3 cats holding a balloon, and 1 mouse holding an umbrella.
       *Q:* Is every cat holding a balloon?
       *A:* No. (pointing to the mouse)

Suppose that a child presented with this scene homes in not on the physical objects it contains but rather on the fact that everyone is holding something. Applying the same reasoning as in the foregoing, such a child (if it failed to get the correct mapping from grammatical form to semantic representation) would arrive at the following interpretation:

(23) [x, y: x holds y] all [: cat(x), balloon(y)]

This says that every individual that holds something is a cat holding a balloon, and thus accounts for the Type-C error in (22).

Appealing though this line of analysis may be, I don't think it is right, for at least two reasons. First, and most importantly, it is at odds with the well-established fact that children are strongly object-oriented in the sense that they tend to concentrate their attention on medium-sized physical objects. Only to give one illustration of this bias, in an experiment conducted by Shipley and Shepperson (1990), 3- to 6-year-olds were presented with toy ducks in various colours, and asked to count the number of different colours, in response to which many children, but especially the younger ones, counted the ducks instead of the colours (see Bloom 2000 for further discussion and references). Results like this are hard to reconcile with the assumption that a child who gives a Type-C response, as in (22), finds the relation of holding more salient than the individuals in this scene. A further reason for distrusting the current proposal is that it doesn't really explain why, apparently, Type-C errors are so much rarer and less persistent than others.

I have the impression that the error in (22) is due to the fact that the child quantifies over all animate individuals in the scene (compare the use of the indefinite pronoun *ones* in (3), which apparently refers to all figures on display, although the target sentence quantifies over circles only). That is to say, the child in (22) seems to obtain something like the following reading:

(24) [x: animal(x)] all [y: cat(x), umbrella(y), x holds y]

If this is right, the problem is that the descriptive material in the sentence doesn't constrain the domain of quantification in any way, be it grammatically or pragmatically. It should be expected, therefore, that children who make Type-C errors have the same mapping problem as children who make Type-A or Type-B errors, but in addition they have a less than secure grasp of the pragmatic principles of interpretation discussed in § 4. This line of explanation seems promising to me, because it does justice to the intuition that Type-C errors are clearly different from, and more serious than, the others, but I must stress once more that in the absence of more systematic data, devising hypotheses tends to be a somewhat gratuitous occupation.


*6. Sundry remarks on earlier accounts*

The chief merit I would claim for my proposal is of course that it gets the main facts right without having to appeal to ad hoc assumptions. But what is nice about it, too, I think, is that it links up to several ideas that have been around in the literature for some time, and allows us to say that there is something right about many alternative accounts. In the present section I will comment on some of these connections.

My analysis locates the problem with universal quantification in one particular area, viz. the syntax/semantics mapping. In this respect, I agree with Donaldson and Lloyd (1974) and Bucci

(1978), among others, who have pointed in the same direction. According to Bucci, for example, problems with a universal sentence arise because semantic relations between words fail to be properly encoded:

> The [universal] sentence is encoded as a simple string or unordered set of substantive words without hierarchical structure […]. This 'structure-neutral' form is initially registered simply as a listing of semantic information including the main content words, e.g., 'all, blue, circles'. Then some further interpretation may or may not be imposed, determined by context or by guessing strategy — not by the original sentence structure as such. (Bucci 1978: 58-59)

It will be clear that my approach is very much in the same spirit. But there are differences, too. For one thing, my account is more explicit than what Bucci proposes. For another, I don't share Bucci's assumption that the semantic representations implicated in children's non-adult responses are 'structure-neutral'. In my view, there is really nothing amiss with these representations as such or with the processes involved; it is just that they are not the appropriate ones for universal sentences. Hence, the process of interpretation is considerably more constrained than Bucci suggests, and this holds not only for its initial phase, but all the way. For, on my account, it is definitely wrong to say that children arrive at an interpretation by mere guesswork: the pragmatic mechanisms that produce Type-A and Type-B responses are fully standard, and only for Type-C responses may it be necessary to relax the assumption that children have adequate mastery of the relevant pragmatic principles.

Another line of approach that bears some resemblance to mine has it that children are prone to misconstrue universal determiners as adverbs of quantification. This is an appealing idea because, in general, adverbials impose relatively weak constraints on their domain of quantification, which makes them more sensitive to pragmatic influences, hence more susceptible to misconstruals of the type I have been discussing. Adverbial theories have been proposed, among others, by Roeper and de Villiers (1993) and Philip (1995); I will briefly discuss the latter here. According to Philip, children tend to construe universal determiners as quantifying over events rather than individuals:[6]

> […] children, for whatever reason, tend to assign to a sentence containing a single determiner universal quantifier [sic] the logical form which adults assign to a sentence containing an adverb of quantification, for which the default mode of quantification is quantification over events. (Philip 1995: 53)

Transposing Philip's analysis into the current framework (and simplifying it somewhat, to boot), the idea is that (25a) is parsed as (25b):

---

[6] Philip doesn't explain why this should be the case, although he tentatively suggests that quantification over events may be more basic than quantification over individuals. This is unlikely, however, in view of children's general object bias (cf. § 5).

(25) a. Every boy is riding an elephant. (= (19a))
     b. [e: …] every [x, y: boy(x), elephant(y), x is riding y in e]

This construal leaves the domain of quantification grossly underdetermined, and calls for further constraints. These are furnished, in Philip's proposal, by a set of special-purpose rules, which might yield like the following, for example:

(25) c. [e, y: elephant(y), y participates in e] all [x: boy(x), x is riding y in e]

(The details of this representation deviate considerably from the official version, but it is the only the principal features of Philip's analysis that I am interested in here.) Broadly speaking, this is similar to what I have proposed: the two accounts agree that a hitch in the grammatical analysis produces a quantifying structure whose domain is virtually unconstrained. The main differences are, first, that according to Philip the child resorts to an entirely different style of quantification, which I find implausible, and secondly, that in Philip's view the mechanisms that fix the domain of quantification are plainly ad hoc.

    Although Drozd's (2001) slogan is the same as mine ('Universal determiners are construed as if they were weak'), the resemblance between our accounts doesn't go much further than that. Drozd explicitly rejects the notion that the locus of the problem lies in the mapping from form to meaning, thereby suggesting that the child's misunderstanding is rather deeper. But if the syntax/semantics mapping is not the problem, what is? Strangely enough, Drozd doesn't say: he doesn't define what it means for a universal determiner to be interpreted as a weak one, which makes his proposal somewhat elusive. (It is rather as if someone were to claim that '+' may be construed as denoting a unary operation without saying which one.) Fortunately, however, Drozd explains at some length what is supposed to follow from his proposal, and it is on his corollaries that I will focus in the following.

    The logical inferences licensed by weak determiners deviate systematically from the inference patterns associated with strong determiners. For example, if D is weak DXY is equivalent to D(X ∩ Y)Y. Thus, 'Some lawyers are crooks' is true iff 'Some lawyers who are crooks, are crooks' is true.[7] Drozd maintains that this equivalence is the source of Type-B errors. If a child interprets *every* as weak and hence intersective, he may infer that (25a) is true iff (26) is, and proceed to verify the latter instead of the former, thus arriving at the conclusion that (25a) is true in a situation in which not all boys ride an elephant.

(26) Every boy who is riding an elephant is riding an elephant.

This proposal raises a number of questions that I shall not go into (though it should be noted that (26) is tautologous, and therefore immune to falsification). Let us turn instead to Drozd's

---

[7] It is true that the latter has a certain air of redundancy about it, but we are interested here in truth-conditional content only.

view on Type-A errors. Given the way Drozd handles Type-B errors, one should expect this class to be treated along the same lines, for if *all* is interpreted as a weak determiner, it must be symmetric as well as intersective, and (25a) must be equivalent not only to (26) but also to:

(27) Every elephant is being ridden by a boy.

A child latching on to *this* equivalence would commit a Type-A error, evidently.[8] However, Drozd doesn't even consider this possibility, and argues that Type-A responses arise in a different way altogether.

What causes Type-A errors, according to Drozd, is the fact that weak determiners are context-sensitive in distinctive ways, and it is this context sensitivity that allows for Type-A construals of universal determiners. The key piece in Drozd's argument is Westerståhl's (1985) example:

(28) Many Scandinavians have won the Nobel prize in literature.

Westerståhl observes that, on its most likely interpretation, (28) means that the number of Scandinavian Nobel prize winners is larger than one would expect, statistically speaking. Drozd's idea is that this somehow carries over to universal sentences, as interpreted by Type-A children:

> As in Westerståhl's account of the preferred in interpretation of [(28)], children who make this error may arrive at a response to the question *Is every boy riding an elephant?* by first comparing the number of elephant-riding boys with what they consider to be the normal or expected frequency of elephant-riders given the situation shown in the picture. If children use the presence of the extra elephant […] to infer that the expected or normal frequency is four [while there are only three boys riding an elephant], they will say *no* […] (Drozd 2001: 358-359)

Note, however, that the parallel Drozd discerns between the standard construal of (28) and the Type-A construal of (25a) is imperfect, at best, because it is hard to see how expected frequencies can interact with the interpretation of *every* in anything remotely like the way they do with the interpretation of *many*. Note, furthermore, that *many* and *few* are in fact unique even among the weak determiners to be sensitive to expected frequencies. So strictly speaking Drozd's thesis is not just that Type-A are caused by weak construals of universal determiners, but rather by a mistaken assimilation of *every* and its kin to the surprise determiners *few* and *many*. Why do children fall into this curious mistake? Drozd doesn't say. However, I don't really want to attack Drozd's proposal. I merely wanted to show that, apart from our *verbatim*

---

[8] Furthermore, Type-C errors should be in for a similar treatment, since it holds for weak determiners that DXY is equivalent to DE(X    Y), where E is the universe of discourse.

agreement about what is the main problem, Drozd's analysis is rather different from mine, and I trust that this much will have become clear.


*Concluding remarks*

The theory presented in this paper is based upon the widely held assumption that children's errors with universal quantification are due to a deficient syntax/semantics mapping. What is new about my approach is that it explains why this mapping should cause trouble, to begin with, and that it provides an explicit account of the comprehension processes underlying children's responses, which also shows how and why pragmatic factors interfere with the interpretation of universal sentences.

Even if this account is correct as far as it goes, it is certainly not the end of the story, if only because we badly need more data. For at present too little is known about Type-C errors, the influence of the collective/distributive distinction (*all* vs. *every* and *each*), and the longitudinal dimension of error patterns, to name only some of the topics that, in my opinion, deserve to be explored further.

*References*

Bloom, P. 2000: *How Children Learn the Meanings of Words*. MIT Press, Cambridge, Mass.

Braine, M.D.S. and B. Rumain 1983: Logical reasoning. In: J.H. Flavell and E.M. Markman (eds.), *Handbook of Child Psychology, Volume 3: Cognitive Development*. Wiley and Sons, New York. Pp. 263-340.

Bucci, W. 1978: The interpretation of universal affirmative propositions. *Cognition* 6: 55-77.

Chien, Y. and K. Wexler 1989: children's knowledge of relative scope in Chinese. *Stanford Child Language Research Forum*.

Crain, S., R. Thornton, C. Boster, L. Conway, D. Lillo-Martin, and E. Woodams 1996: Quantification without qualification. *Language Acquisition* 5: 83-153.

Crain, S. and R. Thornton 1998: *Investigations in Universal Grammar*. MIT Press. Cambridge, Mass.

Donaldson, M. and P. Lloyd 1974: Sentences and situations: Children's judgments of match and mismatch. In: F. Bresson (ed.), *Problèmes Actuels en Psycholinguistique*. Presses Universitaires de France, Paris.

Drozd, K.F. 2001: Children's weak interpretations of universally quantified questions. In: M. Bowerman and S.C. Levinson (eds.), *Language Acquisition and Conceptual Development*. Cambridge University Press. Pp. 340-376.

Drozd, K.F. and W. Philip 1993: Event quantification in preschoolers' comprehension of negation. In: E.V. Clark (ed.)*, Proceeding of the 24th Annual Child Language Research Forum*. Stanford. Pp. 72-86.

Drozd, K.F. and E. van Loosbroek 1999: Weak quantification, plausible dissent, and the development of children's pragmatic knowledge. *Proceedings of the 23rd Annual Boston University Conference on Language Development*, 184-195.

Freeman, N.H., C.G. Sinha, and J.A. Stedmon 1982: All the cars — which cars? From word meaning to discourse analysis. In: M. Beveridge (ed.), *Children Thinking Through Language*. Edward Arnold, London. Pp.52-74.

Freeman, N.H. and J.A. Stedmon 1986: How children deal with natural language quantification. In: I. Kurcz, G.W. Shugar, and J.H. Danks (eds.), *Knowledge and Language*. Elsevier, Amsterdam. Pp. 21-48.

Geurts, B. 1999: *Presuppositions and Pronouns*. Elsevier, Oxford.

Geurts, B. 2000: Review of Crain and R. Thornton (1998). *Linguistics and Philosophy* 23: 523-532.

Geurts, B. and R. van der Sandt, 1999: Domain restriction. In: P. Bosch and R. van der Sandt (eds.), *Focus: Linguistic, Cognitive, and Computational Perspectives*. Cambridge University Press. Pp. 268-292.

Inhelder, B. and J. Piaget 1959: *La Genèse des Structures Logiques Elémentaires: Classifications et Sériations*. Delachaux et Niestlé, Neuchâtel. English translation (1964): *The Early Growth of Logic in the Child: Classification and Seriation*. Routledge and Kegan Paul, London.

Just, M.A. 1974: Comprehending quantified sentences: the relation between sentence-picture and semantic memory verification. *Cognitive Psychology* 6: 216-236.

Meyer, D.E. 1970: On the representation and retrieval of stored semantic information. *Cognitive Psychology* 1: 242-300.

Philip, W. 1992: Distributivity and logical form in the emergence of universal quantification. In: C. Barker and D. Dowty (eds.), *Proceedings of SALT II*. Ohio State University. Pp. 327-345.

Philip, W. 1995: *Event Quantification in the Acquisition of Universal Quantification*. Ph.D. thesis, University of Massachusetts, Amherst.

Philip, W. and M. Takahashi 1991: Quantifier spreading in the acquisition of *every*. T.L. Maxfield and B. Plunkett (eds.), *University of Massachusetts Occasional Papers: Papers in the Acquisition of WH*. GLSA, Amherst, Massachusetts. Pp. 267-282.

Philip, W. and M. Verrips 1994: Dutch preschoolers' *elke*. Paper presented at the 1994 Boston University Conference on Language Development.

Philip, W. and P. Coopmans 1995: Symmetrical interpretation and scope ambiguity in the acquisition of universal quantification in Dutch and English. Ms., University of Utrecht.

Shipley, E.F. and B. Shepperson 1990: Countable entities: developmental changes. *Cognition* 34: 109-136.

Smith, C.L. 1979: Children's understanding of natural language hierarchies. *Journal of Experimental Child Psychology* 27: 437-458.

Smith, C.L. 1980: Quantifiers and question answering in young children. *Journal of Experimental Child Psychology* 30: 191-205.

Takahashi, M. 1991: Children's interpretation of sentences containing *every*. In: T.L. Maxfield and B. Plunkett (eds.), *Papers in the Acquisition of WH*. GLSA, Amherst, Mass. Pp. 303-323.

Westerståhl, D. 1985: Logical constants in quantifier languages. *Linguistics and Philosophy* 8: 387-413.