When "all the five circles" are four: New exercises in domain restriction

Bart Geurts
Bob van Tiel

Nijmegen

Abstract

The domain of a quantifier is determined by a variety of factors, which broadly speaking fall into two types. On the one hand, the context of utterance plays a role: if the focus of attention is on a particular collection of kangaroos, for example, then "Q kangaroos" is likely to range over the individuals in that set. On the other hand, the utterance itself will help to establish the quantificational domain, inter alia through presuppositions triggered within the sentence. In this paper, we concentrate our attention on constructions like the following, in which "the square to which …" is the critical presupposition trigger:

i. Q circles …
ii. Q of these circles …
iii. Q of these five circles …

… have the same colour as the square to which they are connected.

Many theories predict that all instances of these schemata will give rise to the presupposition that every circle is connected to a square. Extending Geurts and van der Sandt's (1999) account, we present an analysis which predicts that these sentences should generally be accepted in a context in which not all the circles are connected to a square, with one exception only: if a quantified sentence is of type (*iii*) and Q is non-intersective, then the sentence should be more likely to be rejected. Furthermore, we predict that manipulating the context so as to make the connected circles more or less salient should have an effect on statements with non-intersective quantifiers only. These predictions were tested in a series of experiments.

August 22, 2014 – to appear in *Topoi*

Introduction

There are five apples on the table. Three pigs enter the room, help themselves to one apple each, and start eating. It would seem obvious that, at this juncture in the narrative, it is true that:

(1) Every pig is eating an apple.

However, for many preschoolers this is anything but obvious, and quite a few children will reject the statement, pointing out, e.g., that: "Those two apples have no pig." This particular example is from Philip and Takahashi (1991); the phenomenon as such was first observed by Inhelder and Piaget (1959), who reported exchanges like the following:

(2) Scene: 14 blue circles, 2 blue squares, 3 red squares.
Experimenter: Are all the circles blue?
Child: No, there are two blue squares.

Problems with quantified sentences are common in preschoolers, and persist at least up to age 7. Not all children have them, but many of them do, and error rates in excess of 50% are not unusual.

As a rule, adults interpret quantified sentences in conformity with the following principle:

> *$\overline{\text{N}}$-constraint*
> A quantifying phrase of the form "Q $\overline{\text{N}}$" must be interpreted with respect to a domain that satisfies the conditions expressed by $\overline{\text{N}}$.

Hence, "most ships" quantifies over ships and only ships, "many shoes" quantifies over shoes and only shoes, and so on. Apparently, the $\overline{\text{N}}$-constraint doesn't come naturally to younger children: they are liable to extend the domain of quantification to entities that are contextually salient but outside the denotation of $\overline{\text{N}}$ (Freeman et al. 1982, Geurts 2003). The $\overline{\text{N}}$-constraint has to be learned, and the learning appears to be hard. Presumably, one of the reasons why this is so is that the domain of a quantifier is determined by multiple factors. On the one hand, the context has its part to play. If the focus of attention is on a collection of ships or shoes, for example, then this raises the likelihood that a quantifier will range over that set. On the other hand, the utterance itself imposes restrictions, both through its truth-conditional content (the $\overline{\text{N}}$-constraint captures part of this) and its presuppositions.

In this paper, we are concerned with the interaction between these factors. In particular, we will study when and how presuppositions triggered within the scope of a quantifier restrict its domain, and how this process interacts with the context:[1]

(3) a. Most people consider their dog an integral family member.
    b. Many dogs know that a seizure is coming.

Clearly, (3a,b) quantify over dog owners and dogs that are about to have a seizure, respectively, so in both cases the quantificational domain is restricted further than is required by the $\overline{\mathrm{N}}$-constraint. In both cases the ultimate cause of the extra restriction is a presupposition, which is triggered by the definite description "their dog" and the factive verb "know", respectively.

There are various accounts of how presuppositions can restrict quantifier domains. In this paper we adopt one that was originally proposed by van der Sandt (1992) and extended by Geurts (1999) and Geurts and van der Sandt (1999). After a general discussion of the semantic framework we adopt (Section 1), we provide a quick review of the key ideas underlying our approach to presupposition in Section 2. Next, in Section 3, we discuss experimental evidence by Chemla (2009) which suggests that sentences like (3a,b) generally give rise to universal presuppositions; for example, in the case of (3a), the preferred construal of the presupposition would be that all people have dogs. However, Chemla's data also indicate that the strength of this preference depends on the quantifier: whereas it is quite strong with universal quantifiers and "no(ne)", it is significantly reduced with quantifiers like "more than/fewer than/exactly *n*".

Since Chemla chose to confine his attention to the interpretative effects of presuppositions, he sought to factor out non-presuppositional determinants of domain restriction by using experimental sentences of the form "Q of the *n* …" By contrast, the focus of our study was on the interaction between the various factors that constrain quantifier domains. In particular, we were interested in the status of the $\overline{\mathrm{N}}$-constraint. Prima facie, the $\overline{\mathrm{N}}$-constraint appears to be absolute, and the logic of Chemla's experimental design requires that it is. But is it really, or is the constraint merely stricter for adults than it is for children?

In our experiments, we presented participants with pictures representing situations that violated the (alleged) universal presupposition of the target sentence, asking whether they considered the sentence true or false in these situations. In Section 4, we extend the DRT analysis to the entire family of sentences used in our experiments, and present evidence

1. Most of our examples are (lightly emended versions of) sentences we found on the internet.

that is in line with our predictions. At the same time, our data contradict Chemla's finding that the quantifier "no(ne)" patterns with the universal quantifiers. Finally, in Section 5, we show that participants' intuitions about quantified sentences are affected by the way visual information is displayed, and that the $\overline{\text{N}}$-constraint is not always obeyed by adults, either: in some contexts, sentences of the form "Each of these *n* A B", where B contains a presupposition trigger, are more often accepted than rejected when the number of AB-individuals is smaller than *n*.

1. Discourse representation theory

The theoretical framework adopted in this paper is that of dynamic semantics, and more specifically Discourse Representation Theory (DRT; Kamp 1981, Kamp and Reyle 1993). In this section, we provide a thumbnail sketch of DRT.

A discourse representation structure (DRS) is a representation built up as the discourse unfolds, and it consists of two parts: a universe of discourse referents (which are "introduced" by the DRS) and a set of DRS-conditions which encode the information that has accumulated on these discourse referents. The following DRS represents the information that there are two individuals, one of which is called "Pedro" while the other is a donkey, and that the former was chasing the latter:

(4)

x y
Pedro x
donkey y
x was chasing y

The universe of this DRS contains two discourse referents, x and y, and its condition set is {Pedro x, donkey y, x was chasing y}.

DRSs are given a model-theoretic interpretation by means of embedding functions, which are partial functions from discourse referents to individuals in a given model M. An embedding function *f* verifies (4) in M iff *f*'s domain contains at least x and y, and in M it is the case that *f*(x) is called "Pedro", *f*(y) is a donkey, and *f*(x) was chasing *f*(y).

The DRS in (4) is designed to reflect the intuitive meaning of:

(5) Pedro was chasing a donkey.

In the absence of contextual information, the semantic representation of (5) is (4). So the indefinite expression "a donkey" is not treated as a regular quantifier; rather, it prompts the introduction of a new discourse referent, y, and contributes the information that y a donkey.

If a discourse opens with an utterance of (5), the DRS in (4) is constructed, and this DRS forms the background against which the next utterance is interpreted, which might be (6a), for example:

(6) a. He caught it.

b.

v w
v caught *w*

(6b) is the DRS that reflects the semantic content of (6a) before the pronouns are resolved. In this DRS, the anaphoric pronouns "he" and "it" in (6a) are represented by the discourse referents *v* and *w*, respectively, which are italicised to indicate that they want to be identified with discourse referents that are given already. (6a) is uttered in the context of (4), so the next step in the interpretation of this sentence is to merge the DRS in (6b) with that in (4), the result of which is (7a):

(7) a.

x y *v w*
Pedro x donkey y x was chasing y *v* caught *w*

b.

x y
Pedro x donkey y x was chasing y x caught y

Since (6a) is immediately preceded by (5), the most likely antecedents of "he" and "it" are "a farmer" and "a donkey", respectively. At DRS level, this is represented by equating *v* with x and *w* with y, which results in (7b). This DRS is verified in any model featuring an individual called "Pedro" who chased and caught a donkey.

Thus far, we have considered DRSs with simple conditions, but DRT comes into its own when complex conditions are called for:

(8) a. Pedro doesn't have a donkey.

b.

x y
Pedro x
¬ [y | donkey y, x owns y]

(8b) is the sentence DRS corresponding to (8a). This DRS contains a condition that consists of a DRS prefixed by a negation sign. An embedding function *f* verifies (8b) in a model M iff *f*(x) is called "Pedro" and *f* cannot be extended to a function *g* which verifies the embedded DRS; that is to say, no such *g* should map y onto a donkey owned by Pedro.

The negated DRS in (8b) contains a token of the discourse referent x which is introduced in the main DRS. The embedded DRS introduces an additional discourse referent, y, which is associated with the indefinite NP "a donkey", and whose scope is delimited by the sub-DRS in which it is introduced. This explains the observation that if (8a) were followed by (9a), for example, the pronoun could not be linked to the indefinite:

(9) a. It is grey.

b.

x *v*
Pedro x
v is grey
¬ [y
donkey y
x owns y]

Whenever one DRS is embedded in another, the latter is accessible from the former, but not the other way round. Hence, in (9b), the discourse referent *v*, which represents the neuter pronoun in (9a), does not have access to y, because y is introduced in a DRS that is not accessible to the DRS in which *v* is introduced, and therefore it is not possible to bind *v* to y. Note that accessibility, thus understood, follows from the model-theoretic interpretation of DRSs.

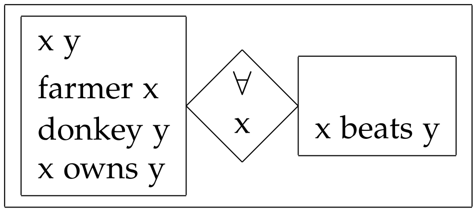
The following example illustrates the DRT treatment of quantified sentences:

(10) a. Every farmer who owns a donkey$_i$ beats it$_i$.

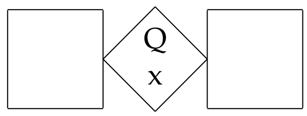
b.

[x y
farmer x
donkey y
x owns y] ⟨∀ x⟩ [*v*
x beats *v*]

c.



There are two ways of spelling out the interpretation of so-called “duplex conditions” of the form:



On one construal, an embedding function *f* verifies the duplex condition in (10c) iff the following holds for every individual *i:*

> If any *g* extends *f* in such a way that *g*(x) = *i* and *g*(x) is a farmer who owns a donkey, *g*(y), then *g*(x) beats *g*(y).

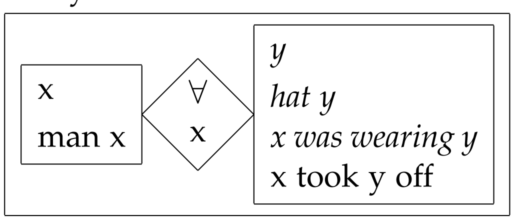
Thus construed, (10) entails that every donkey-owning farmer beats all his donkeys. The second interpretation is weaker: it says that every donkey-owning farmer beats at least one of his donkeys. For our current purposes, this ambiguity is irrelevant. The only thing that matters is that either interpretation makes the restrictor DRS on the left accessible to the scopal DRS on the right. Therefore, the discourse referent y in (10b) is accessible to v, and the latter may be equated to the former, which yields (10c).

2. *Accommodation*

By most accounts, including our own, presuppositions are information that is presented as given. For example, the proper function of a definite description is to refer to a given individual; this is what distinguishes definites from indefinites. In the theory adopted in this paper, anaphora and presupposition are treated as closely related phenomena: we treat the former as a special instance of the latter. A presupposition is a chunk of information which is introduced into a DRS in a position that is determined by the sentence grammar, and which is presented as given in that position. Givenness here simply means that the information in question is available in a DRS that is accessible from the position in which the presupposition is introduced. Thus the treatment of anaphoric pronouns outlined in the foregoing is extended to (other) presupposition triggers.

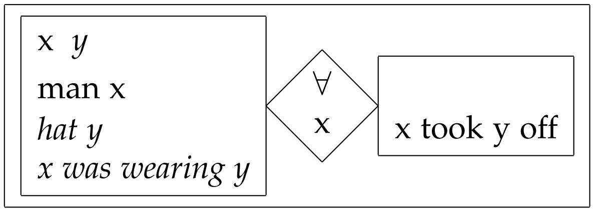
To say that presupposed information is presented as given is to imply that it doesn't have to *be* given (Karttunen 1974, Stalnaker 1974). For example, definite descriptions are often used to introduce information that is strictly speaking new. If I mention "my bicycle" to someone who didn't know that I'm a bicycle owner, I'm pretending it is already common ground that I am. My addressee will normally be prepared to go along with my pretence because she realises that, in order to avoid using a presuppositional expression, I would have had to assert that I own a bicycle, which would have been unnecessarily prolix, given that it is a thoroughly unsurprising fact. This is what, due to Lewis (1979), is known as "accommodation" of presupposed information.

If a presupposition *p* is interpreted by way of accommodation, it is added to the context in which *p* was triggered. This seems straightforward enough. However, in a dynamic framework like ours, there will generally be more than one context in which *p* might be accommodated. In DRT, these contexts are represented by boxes and boxes within boxes.

(11) a. Every man took his hat off.
b.


In (11b), which is the DRS associated with (11a), the definite description "his hat" has triggered a presupposition (marked in italics) in the scope of the quantifier. In principle, this presupposition could be accommodated in any of the three sub-DRSs in (11b): the main DRS, the sub-DRS representing the quantifier's restrictor, and the sub-DRS representing its scope, which is where the presupposition was triggered to begin with. These three options are known as "global", "intermediate", and "local" accommodation, respectively. In this case, as indeed in all the cases we will be concerned with, global accommodation is ruled out, because the presupposition contains a discourse referent, x, which is introduced in the restrictor DRS, and would become free if the presupposition were accommodated in the main DRS. This leaves us with two possibilities. If the presupposition is accommodated locally, it remains where it originated and (11a) is read in effect as, "Every man was wearing a hat and took it off"; hence, on this reading, the presupposition gives rise to the universal inference that every man was wearing a hat. Intermediate accommodation results in the following DRS:

(12)



This says that every man with a hat took it off, or in other words: intermediate accommodation causes the quantifier's domain to become restricted to individuals with hats.

So, we have two readings for (11a), both of which seem to be plausible. How are we going to choose? In our general account of presupposition, it is assumed that, ceteris paribus, there is a preference for accommodating presuppositions as globally as possible; this is an instance of what Geurts (2000) calls the "Buoyancy Principle", and that is the term we will use, too.[2] Hence, we predict that, all things being equal, intermediate accommodation is preferred to local accommodation. However, in the following, the Buoyancy Principle and the predictions it licenses will play a very minor role. In particular, the explanation we will offer for our own experimental data does not hinge on it.

3. *All quantifiers are not alike*

Chemla (2009) reports on two experiments designed to probe speakers' intuitions about sentences like (11a), i.e. quantified statements whose nuclear scope contains a presuppositional expression. In the following, we will confine our attention to what we take to be the main contribution of Chemla's study, which concerns the question of whether such sentences give rise to universal inferences across the board, as predicted by Heim (1983) and Schlenker (2008), among others.

In his first experiment, Chemla used an inference paradigm with items like the one shown in Fig. 1, each item presenting a one-step inference from a target sentence to a candidate presupposition. Participants were asked to indicate whether or not the premiss suggested the conclusion.

2. Although the Buoyancy Principle is *de facto* accepted by most dynamic theories of presupposition, its import is partly dependent on the details of the theory. In particular, in non-representational versions of dynamic semantics, intermediate accommodation is not an option for technical reasons, which reduces the Buoyancy Principle to a preference for global as opposed to local accommodation.

"None of these 10 students knows that he is lucky"

suggests that:

Each of these 10 students is lucky.

NO YES

Figure 1: Sample trial from Chemla (2009, Experiment 1).

The quantifiers featuring in the materials were "each", "no", "more than 3", "fewer than 3", and "exactly 3". Chemla's main finding was that, contrary to what Heim, Schlenker, and others had predicted, presuppositional expressions in the nuclear scope of quantifiers did not give rise to universal inferences indiscriminately: whereas sentences with "each" and "no" prompted universal inferences between 80% and 90% of the time, the corresponding rates for "more than 3", "fewer than 3", and "exactly 3" ranged between 50% and 60%. Chemla's second experiment, in which participants were asked for graded rather than dichotomous judgments, corroborated these results.

Chemla's data raise two main issues. First, it appears that speakers' intuitions about the target sentences were mixed, and more so in some cases than in others. At a general level of description, this is not unexpected if we view the matter in terms of domain restriction, considering that domain restriction results from an interaction between multiple factors, of which presupposition is only one. Still, it remains to be seen how this interaction plays out; which is our main concern in the following.

Secondly, it is curious that the quantifier "no" should pattern with "each" rather than with the intersective quantifiers.[3] Given that, from a semantical point of view, "no" is intersective, we should ask why, in Chemla's study, it failed to behave like the other intersective quantifiers. This is one of the reasons why we wanted to see if Chemla's findings would replicate in a different experimental paradigm.

Another reason had to do with domain restriction. Chemla tried to

3. Intersective quantifiers are those for which the truth conditions of a sentence of the form "Q AB" can be defined solely in terms of the cardinality of the intersection between the extensions of A and B. For example, while "some", "more than two", and "no" are intersective, universal quantifiers and proportional quantifiers ("most", "more than 33%") are not.

control for domain restriction by only using quantifiers with numerical restrictors, like "each of these 10 students", supposing that participants would honour the $\overline{\text{N}}$-constraint and fix the domain of quantification to 10 students. Although this assumption seems reasonable, we don't know for a fact that it is justified, and in view of our opening discussion it is possible that it might not be. More generally, whereas Chemla's strategy was to factor out non-presuppositional determinants of domain restriction, we wanted to study how presuppositional and non-presuppositional factors interact to determine the domain of quantification. Hence, we decided to go over the same ground as Chemla, but with a different method.

In our experiments, we used a verification paradigm: on each trial, participants were presented with a picture and a quantified sentence with a presupposition trigger in its scope, and were asked whether the sentence was true of the depicted situation. In all experimental trials, the picture verified the standard truth-conditional interpretation of the target sentence, but *not* the universal presupposition predicted by Heim, Schlenker, and others. This set-up allowed us to gauge the robustness of the universal inferences observed by Chemla. More importantly, however, in this way we could also study the interaction between linguistic and contextual factors: since the visual display was the most conspicuous part of the extra-linguistic context, we could manipulate that and check for effects on participants' responses.

4. *Analysis, predictions, and first findings*

Table 1 lists the quantifiers we used in our study. From right to left, we have the complex quantifiers studied by Chemla, which we call "DN-partitives" (D for "demonstrative", N for "numerical"), "D-partitives", and "plain quantifiers". The target sentences we used in our experiments were all of the following form, where Q ranged over the quantifiers in Table 1:

(S) Q {circle has/circles have} the same colour as the square to which {it is/they are} connected.

In all experimental trials, the picture showed Q circles which had the same colour as a square they were connected to, but at least one circle wasn't connected to a square.

In this section, we first present our analysis of this family of sentences, and then report on our experiments. We will use "S[every]" to refer to the sentence, "Every circle has the same colour as the square to which it is connected", and similarly for the other quantifiers.

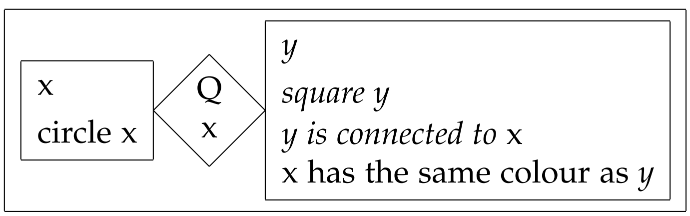
Plain Qs	D-partitive Qs	DN-partitive Qs
every	each of these	each of these five
no	none of these	none of these five
exactly two	exactly two of these	exactly two of these five
more than two	more than two of these	more than two of these five
less than three	less than three of these	less than three of these five

Table 1: Quantifiers used in Experiments 1 and 3.

Plain quantifiers

Further to the discussion of Section 2, up to resolution of the key presupposition, semantic representations for sentences with plain quantifiers look like this:

(13)



As discussed in Section 2, the presupposition can be accommodated either in Q's restrictor or its scope (where it was triggered). For sentences with intersective quantifiers, including S[no], the resulting truth conditions are the same in either case, and presupposition projection will never give rise to universal inferences.

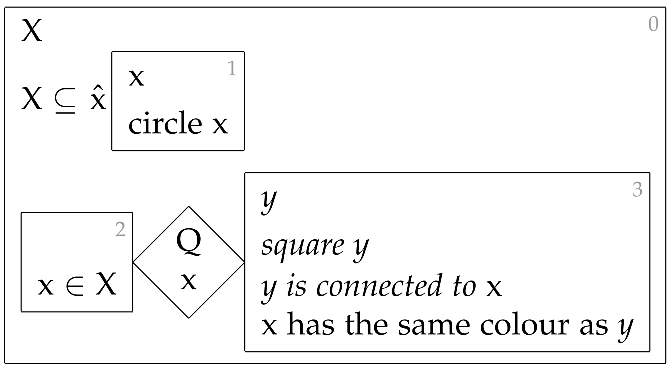
If Q is non-intersective, it does make a difference where the presupposition is accommodated. If it is accommodated in the restrictor, S[every] is interpreted as, "Every circle that is connected to a square has the same colour as it"; if it is accommodated in the scope of "every", the presupposition gives rise to a universal inference: "Every circle is connected to a square and has the same colour as it." Since, in the critical items of our study, S[every] was false on the second construal, and we assume that hearers are charitable and, ceteris paribus, prefer true readings to false ones, we predict a preference for intermediate accommodation, and thus for the weaker reading. That is to say, we forecast that, ceteris paribus, S[every] may be judged true even if not every circle is connected to a square.

In short, if Q is a plain quantifier, be it intersective or not, we predict that S[Q] does not require a context in which all circles are connected to a square.

D-partitive quantifiers

To a first approximation, sentences with D-partive quantifiers, like S[each of these], can be represented as follows:

(14)



We assume that the partitive prompts the introduction of a plural discourse referent, X, which represents a set of circles, and that Q ranges over that set. In actual usage, X might be further restricted by the context, notably by pointing, but since there was no pointing in our experiments, we can leave such constraints out of account.

In (14), DRSs are labeled by numbers; we will refer to the main DRS as C_0, and to its sub-DRSs as C_1, C_2, and C_3. Since X denotes a subset of $\hat{x}\,C_1$, i.e. the set of all circles, and X is Q's domain, C_1 is accessible to C_2, which means that the presupposition triggered in C_3 can be accommodated in C_1, C_2, or C_3.[4] (As before, global accommodation is not an option.)

It will not be hard to see that the extra structure that distinguishes (14) from (13) is truth-conditionally inert, nor does it make any difference to the predictions we derive concerning the interpretation of the presupposition triggered in C_3. If Q is a intersective quantifier, it doesn't matter whether the presupposition is accommodated in C_1, C_2, or C_3; truth-conditional content will be the same in any case. If Q is a non-intersective quantifier, accommodation in C_1 and C_2 will yield the same truth conditions, which are weaker than when the presupposition is accommodated in C_3. Only in the last case will the presupposition give rise to a universal inference, so ceteris paribus, S[each of these] may be judged true even if

4. In Kamp and Reyle's (1993) version of DRT, C_1 would not be accessible to C_2. There are several ways of correcting this; see Geurts (2012) for discussion.

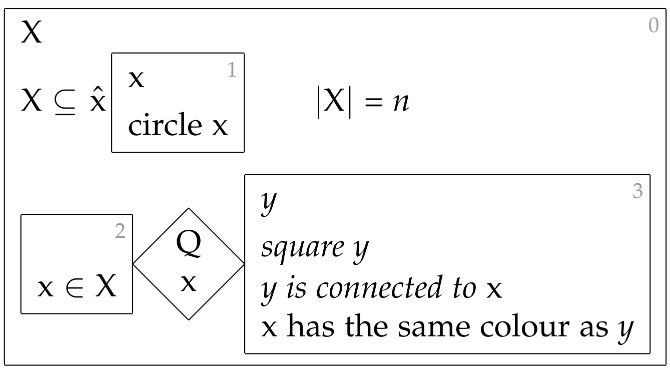
not every circle is connected to a square.

The upshot of these observations is that, for our current purposes, there are no relevant differences between plain and D-partitive quantifiers. Most importantly, as in the case of plain quantifiers, we predict that if Q is D-partitive, be it intersective or not, S[Q] does not require a context in which all circles are connected to a circle.

DN-partitive quantifiers

Whereas we expect sentences with plain and D-partitive quantifiers to be interpreted alike, our account predicts that sentences with DN-partitives, like S[each of these *n*], should behave differently. As one might expect, our analysis of such sentences looks very much like (14), the only difference being that it imposes a constraint on X's cardinality:

(15)



This small addition has notable implications for the interpretation of the presupposition triggered in the scope of Q:

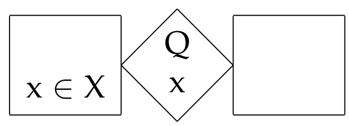
i. Accommodation in C_1 entails that all *n* circles in X are connected to a square. The resulting reading will be at least as strong as any other, regardless of whether Q is intersective or not.

ii. If Q is intersective, accommodation in C_2 produces the same truth conditions as accommodation in C_3, which are weaker than when the presupposition is accommodated in C_1.

iii. If Q is non-intersective, accommodation in C_3 produces the same truth conditions as accommodation in C_1, which are stronger than when the presupposition is accommodated in C_2.

Now suppose that S[Q] is uttered in a situation in which the universal presupposition is false: fewer than *n* circles are connected to a square. Hence, accommodation in C_1 will make the sentence come out false. Assuming the hearer is charitable, she will try to save the assumption that the speaker intends to make a true statement, which can be done by accommodating the presupposition in C_2 or C_3. If Q is intersective, this choice is truth-conditionally inert, but if Q is non-intersective, accommodation in C_3 yields the same truth conditions as accommodation in C_1, so the only way of mitigating the effect of the presupposition is by accommodating it in C_2.

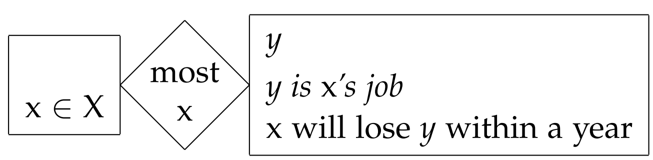
However, there is reason to suppose that, in a configuration like (15), accommodation in C_2 is dispreferred. (The same goes for (14), but there nothing hinges on the choice between C_2 and either of the other sub-DRSs.) The duplex condition in (15) is of the form,



and is a purely formal one in the sense that it serves merely to distribute over X. In (15), all the descriptive content associated with the restrictor of the quantifier is in C_1, and therefore it makes sense to assume that any further restrictions on the domain of quantification are preferentially added to C_1 rather than C_2. Hence, we hypothesise that hearers will be reluctant to add material conditions to C_2 by way of accommodation. This reluctance also explains the following contrast:

(16) In this region, nearly all men over fifty are jobless, and …
  a. ?most of them will lose their jobs within a year.
  b. most of them who aren't will lose their jobs within a year.

The DRS representing (16a) will contain a condition of the form,



where X represents a set of mostly jobless men. If the presupposition triggered by "their jobs" could be accommodated in the restrictor of the quantifier, (16a) should have the same interpretation as (16b) and be no less felicitous (cf. Beaver 2001). However, (16a) is clearly odd, thus confirming our hypothesis that accommodation in purely distributive restrictor DRSs is dispreferred.

Assuming, then, that this hypothesis is correct, we should expect that, in a situation in which the universal presupposition is false, hearers will be more likely to accept S[Q of these *n*] if Q is intersective than if it isn't.

For, whereas in the former case, the effect of the presupposition can be weakened by accommodating it in C_3, in the latter this can only be done by accommodating the presupposition in C_2, which is dispreferred.

Experiment 1

In the foregoing we analysed quantified sentences of the following form, in which a presupposition is triggered in the scope of the quantifier:

(S) Q {circle has/circles have} the same colour as the square to which {it is/they are} connected.

Our analysis predicts that such sentences should be generally acceptable in contexts that falsify the strong construal of the presupposition (i.e., each circle is connected to a square), with one exception: if Q is "each of these *n*", then the sentence should be rejected more often than for the other quantifiers. In order to test this prediction, we presented participants with items like the one shown in Fig. 2.[5] In each item only four of the five circles were connected to an adjacent square, hence the universal presupposition was false in all cases. However, in all target items, Q circles were connected to a square of the same colour. For each of the quantifiers listed in Table 1, five target items were constructed, varying the distribution of the colours, the connections between squares and circles, and the position of the unconnected circle.

The response pattern we observed confirmed our predictions. For intersective quantifiers, the rates of "true" responses exceeded 85% across the board. The same held for universal quantifiers in the Plain and D-partitive conditions, but not in the DN-partitive condition, where only 11% of the responses were positive. These data are in line with the prediction that our target sentences should be accepted in practically all cases, with one exception only: sentences of the form "Each of these *n* …"

Although, prima facie, these results contradict Chemla's flat out, the tension is more apparent than real, simply because we didn't use the same experimental paradigm. Recall that Chemla, using an inference paradigm, found that participants endorsed high rates of universal presuppositions for sentences with DN-partitives. This is what we would predict, too, courtesy of the Buoyancy Principle (Section 2). Referring to our own example in (16), that principle says that hearers should prefer to accommodate the presupposition in C_1, which produces the universal inferences

5. For more detailed presentations of this and the following experiments, we refer to the Appendix.

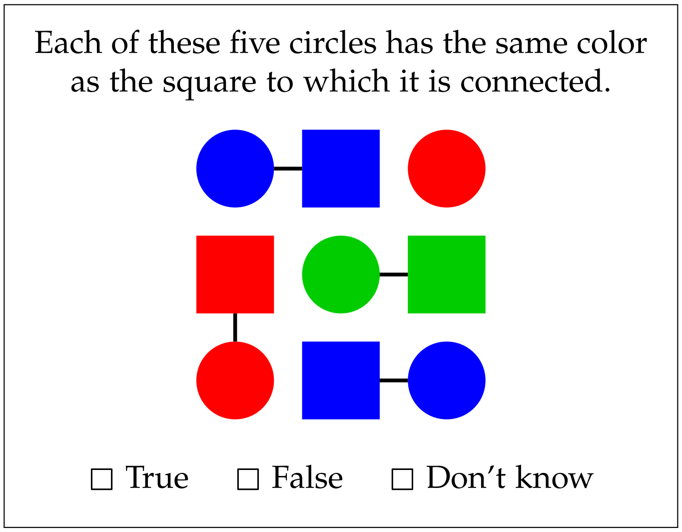


Figure 2: Item used in Experiment 1.

Chemla observed in the majority of cases. It is only when the context makes the universal inference false, as it did in our experiment, that hearers will try to find an alternative way of accommodating the presupposition. What remains unexplained, however, is why the tendency to draw the universal inference should be weaker in some cases than in others, as Chemla's data indicate.

Another puzzle concerns the quantifier "no(ne)", which in Chemla's study behaved like "each", whereas in ours it went with the other intersective quantifiers. Our findings are in line with naturally occurring examples like the following:

(17) To my knowledge none of these bands ever had *their Casio cassette player* eat *the lip synch tape* in the middle of a live performance.

(18) Timo: What was the first Star Trek episode in which our heroic captain was "James T. Kirk" instead of just "James Kirk" (or "James R. Kirk")?

Christopher: The idea that Kirk *had* a middle initial was obviously "conceived" in the second pilot, and it's so common for people to have middle initials that it's a rather bizarre question to ask.

Starbreaker: It's not a bizarre question. None of the other characters ever had *their middle initial* featured so prominently.

According to our intuitions, neither (17) nor Starbreaker's comment in (18) is likely to give rise to a universal presupposition. That is, (17) was probably not meant to suggest that all the bands under discussion use a Casio cassette player and a lip synch tape, nor does Starbreaker's second sentence imply that all Star Trek characters have an odd number of initials. By contrast, the following sentences do carry these implications:

(19) a. Each of these bands at some time had their Casio cassette player eat the lip synch tape in the middle of a live performance.
b. All the other characters had their middle initial featured just as prominently as Kirk.

These observations suggest that universal presuppositions may be less robustly associated with "no(ne)" than with "each" and its kin. However, even if this is true, it remains to be seen why in Chemla's study "no(ne)" failed to pattern with the other intersective quantifiers. We don't have an answer to that question, but we will present additional evidence that, in a verification paradigm, "no(ne)" behaves like a true-blue intersective quantifier.

5. The power of the picture

In Experiment 1 we found, as predicted, only one case in which falsity of the universal presupposition caused most participants to reject a quantified sentence: this happened when participants were presented with items like the one shown in Fig. 2, which featured the target statement:

(20) Each of these five circles has the same colour as the square to which it is connected.

This was a robust effect in the sense that (20) elicited negative responses 87% of the time, which indicates that the $\overline{\text{N}}$-constraint is quite strong, even if these items were accepted in a substantial number of cases, as well. We wanted to know how strong the $\overline{\text{N}}$-constraint really is when it is pitted against other determinants of domain restriction. As discussed in the introduction to this paper, quantifier domains are constrained by contextual as well as linguistic factors. Since in our experimental paradigm the picture is the most conspicuous part of the context, we set out to determine whether response patterns could be influenced by manipulating the layout of the visual display. The idea was that by making the set of connected circles more salient, more participants might be tempted to accept the critical sentence, despite the $\overline{\text{N}}$-constraint.

This was the primary goal of our follow-up experiment. Its secondary goal was to replicate our findings for the quantifier "no(ne)", since that was the main point in which our data deviated from Chemla's.

Experiment 2

But first we wanted to see whether, independently from presuppositional constraints, domain restriction can be influenced by manipulating the picture in the first place. We did this by asking two groups of participants to judge whether the sentence, "All the squares are red", is true of the pictures in Fig. 3. The two pictures were identical, except that in the A-picture fewer of the squares were red than in the B-picture. Still, whereas the A-group accepted the target statement 52% of the time, the B-group did so only 16% of the time.

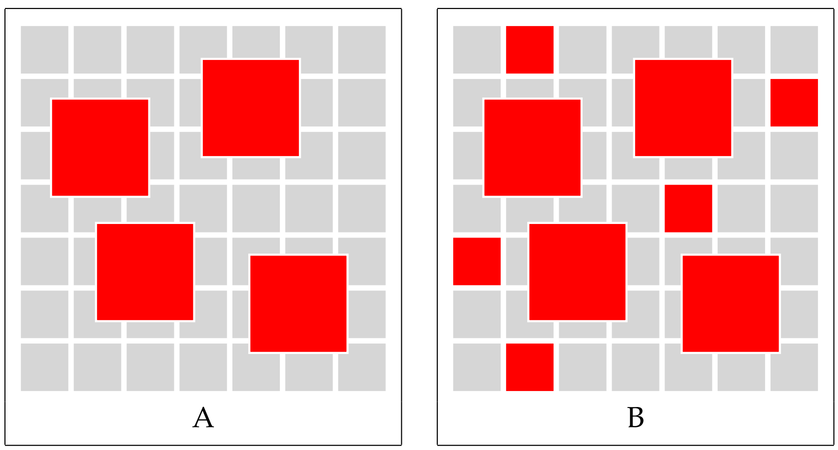


Figure 3: Pictures used in Experiment 2.

This result may not seem very surprising, but it illustrates an important point: in the verification task, the picture can affect the domain of quantification, and thus the interpretation of the target sentence (cf. Rubin 1921, Drozd and van Loosbroek 1998, Philip and Lynch 2000, Pietroski et al. 2009). The A-picture invites the viewer to separate the squares into two groups: red ones in the foreground, grey ones in the background. In the B-picture, visual grouping is not as easy, because the small red squares blur the foreground/background distinction, making it less inviting to treat the large squares as a distinct group.

Note that the effect of visual grouping is a matter of degree: on the A-item, opinions were more or less evenly divided, and although most participants rejected the target sentence in the B-condition, it was still accepted by a substantial minority. Our next question, and indeed the main question of this paper, is whether the same holds of the $\overline{\text{N}}$-constraint, or whether this constraint is absolute.

Experiment 3

In our final experiment, we manipulated pictures so as to facilitate visual grouping of the circles that were connected with a square (Figs. 4 and 5). We adopted three measures to achieve this effect:

i. *Use fewer colours:*
One way of raising the salience of the key information in a picture is by making it simpler, and one way of doing that is by using fewer colours. Hence, while in Experiment 1, all pictures were in three colours, in the current experiment we included two-colour items, which were matched by three-colour controls (Fig. 4A vs. Fig. 4B).

ii. *Simplify the layout of the pictures:*
Another way of making a picture easier to read is by avoiding clutter and presenting circles and squares in a more orderly fashion than in the first experiment. We did this by including a condition in which pairs of circles and squares neatly lined up in a row (Fig. 5).

iii. *Reduce the number of connected circles:*
Finally, we thought the connected circles might become more salient if there were fewer of them, and therefore we included items in which the number of connected circles was just enough to justify the use of a plural, that is to say, two (Fig. 5).

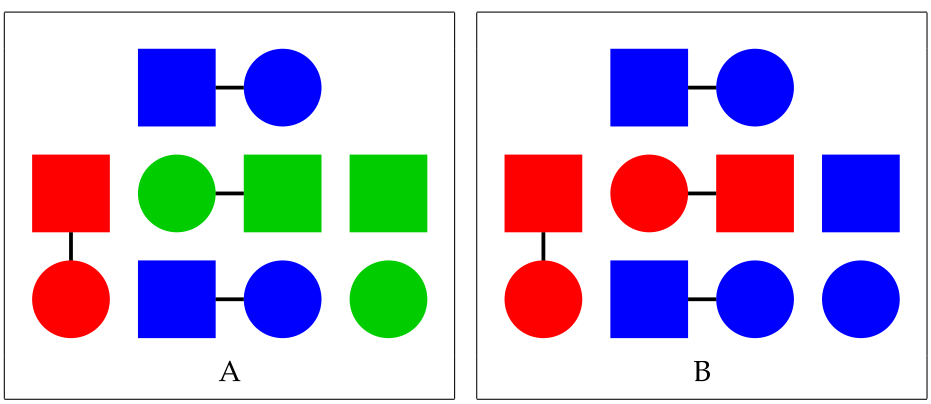


Figure 4: Pictures used in Experiment 3.

While the purpose of these manipulations was to raise the salience of the connected circles, Experiment 3 introduced a further change with respect to Experiment 1, which served to rule out a potential confound. Based on our informants' comments, we received the impression that a small number of them might have been influenced in their judgment by the fact that the unconnected circle in Fig. 2 "has no square to connect to". Strictly speaking, this is false, but since there were no items containing figures with more than one connection, the observation makes sense, and as it seemed to strengthen participants' tendency to reject the critical items, we chose to balance the circles and the squares in such a way that every unconnected circle was adjacent to an unconnected square.

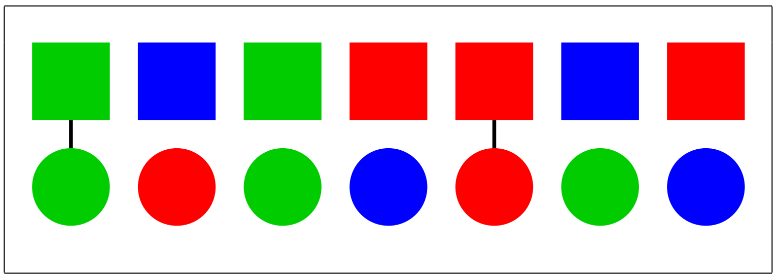


Figure 5: Picture used in Experiment 3.

In Experiment 3, sentences of the form S[each of these n] and S[none of these n], were presented with pictures like the ones shown in Figs. 4 and 5; in all cases, n was the total number of circles in the display. Three groups of participants either saw 3×4 three-colour pictures (Fig. 4A), 3×4 two-colour pictures (Fig. 4B), or 2×7 three-colour pictures (Fig. 5). Table 2 gives the rates of positive responses for each of the conditions, as well as the corresponding rates observed in Experiment 1. Across conditions, the rates at which "none" sentences were accepted were reliably different from their universal counterparts, but not from each other. Rates for sentences with "each" were reliably different between all conditions, except for 3×4 two-colour vs. 2×7 three-colour pictures (last two lines in Table 2).

PICTURE DIMENSION	NUMBER OF COLOURS	*each of these n*	*none of these n*	
3×3	3	13	95	Exp. 1, Fig. 2
3×4	3	35	100	Exp. 3, Fig. 4A
3×4	2	68	92	Exp. 3, Fig. 4B
2×7	3	65	96	Exp. 3, Fig. 5

Table 2: Percentages of "true" responses to items with "each/none of these n" in Experiments 1 and 3.

First off, these results replicate the finding of Experiment 1 that, in situations that falsify the universal presupposition, sentences with "none" are accepted more often than sentences with "each": whereas the acceptance rates for "none" sentences were at or near ceiling level throughout, the corresponding rates for "each" sentences fluctuated between 13% and 68%. This contrasts with Chemla's finding that, in an inference paradigm, participants endorse universal presuppositions at equal rates for both types of sentence. We don't have a solution to this puzzle, but the data of the last experiment confirm that it *is* a puzzle.

Secondly, our decision to balance the circles and the squares seems to have had positive effect on the rates at which universal sentences were accepted: whereas S[each of these five] was accepted 13% of the time with pictures like the one shown in Fig. 2, the corresponding rate for the category of pictures exemplified by Fig. 4A was 35%.

Finally, and most importantly, the outcome of this experiment shows that, even with sentences of the form "Each of these *n* A B", where B contains a presupposition trigger, the domain of quantification can be restricted by contextual factors: in two of the three conditions, sentences of this type were more often accepted than rejected when the number of AB-individuals was smaller than *n*. Most strikingly, perhaps, "Each of these *seven* circles has the same colour as the square to which it is connected" was judged true 68% of the time in a situation in which only *two* circles were connected to a square. Our analysis explains how this is possible. Hearers who accept this sentence even when only two circles are connected must interpret it such a way that the presupposition is accommodated in a position that is otherwise dispreferred, which requires that the context support this restriction. In the materials of our experiment, it was the visual context that provided the necessary support.

Conclusion

In the introduction to this paper, we discussed the $\overline{\text{N}}$-constraint, which states that a quantifying phrase of the form "Q $\overline{\text{N}}$" is to be interpreted with respect to a domain that satisfies the conditions expressed by $\overline{\text{N}}$. We saw that children's interpretations of universal sentences sometimes violate this constraint: if the cows are less salient than the cowsheds, a three-year old is liable to construe "all the cows" as ranging over cowsheds rather than cows (Freeman et al. 1982). Presumably, such blatant violations of the $\overline{\text{N}}$-constraint don't occur in adults, but in the foregoing we have shown that grown-ups do not always respect the constraint, either. This observation lends further support to the view that quantifier domains are determined by a variety of factors, and that when these factors contradict one another, it is not necessarily the linguistic ones that prevail.

Acknowledgment

This research was supported by a grant from the Netherlands Organisation for Scientific Research (NWO).

References

Beaver, D. (2001). *Presupposition and assertion in dynamic semantics*. Stanford: CSLI Publications.

Chemla, E. (2009). Presuppositions of quantified sentences: experimental data. *Natural language semantics 17*, 299–340.

Drozd, K. and E. van Loosbroek (1998). Weak quantification, plausible dissent, and the development of children's pragmatic competence. In *Proceedings of the 23rd Boston University conference on language development*, pp. 184–195. Somerville, Massachusetts: Cascadilla Press.

Freeman, N., C. Sinha, and J. Stedmon (1982). All the cars—which cars? From word meaning to discourse analysis. In M. Beveridge (Ed.), *Children thinking through language*, pp. 52–74. London: Edward Arnold.

Geurts, B. (1999). *Presuppositions and pronouns*. Oxford: Elsevier.

Geurts, B. (2000). Buoyancy and strength. *Journal of semantics 17*, 315–333.

Geurts, B. (2003). Quantifying kids. *Language acquisition 11*, 197–218.

Geurts, B. (2012). Accessibility and anaphora. In K. von Heusinger, C. Maienborn, and P. Portner (Eds.), *Semantics: an international handbook of natural language meaning, volume 2*. Berlin: De Gruyter Mouton.

Geurts, B. and R. A. van der Sandt (1999). Domain restriction. In P. Bosch and R. A. van der Sandt (Eds.), *Focus: linguistic, cognitive, and computational perspectives*, pp. 268–292. Cambridge University Press.

Heim, I. (1983). On the projection problem for presuppositions. In M. Barlow, D. P. Flickinger, and M. T. Wescoat (Eds.), *Proceedings of the second West coast conference on formal linguistics*, pp. 114–125. Stanford Linguistics Association.

Inhelder, B. and J. Piaget (1959). *La genèse des structures logiques élementaires: classifications et sériations*. Neuchâtel: Delachaux et Niestlé.

Kamp, H. (1981). A theory of truth and semantic representation. In J. Groenendijk, T. Janssen, and M. Stokhof (Eds.), *Formal methods in the study of language*, pp. 277–322. Amsterdam: Mathematical Centre Tracts 135.

Kamp, H. and U. Reyle (1993). *From discourse to logic*. Dordrecht: Kluwer.

Karttunen, L. (1974). Presupposition and linguistic context. *Theoretical linguistics 1*, 181–194.
Lewis, D. K. (1979). Scorekeeping in a language game. *Journal of philosophical logic 8*, 339–359.
Philip, W. and E. Lynch (2000). Felicity, relevance, and the acquisition of the grammar of *every* and *only*. In *Proceedings of the 24th Boston University conference on language development*, pp. 583–596. Somerville, Massachusetts: Cascadilla Press.
Philip, W. and M. Takahashi (1991). Quantifier spreading in the acquisition of *every*. In T. Maxfield and B. Plunkett (Eds.), *Papers in the acquisition of WH*, pp. 267–282. GLSA.
Pietroski, P., J. Lidz, T. Hunter, and J. Halberda (2009). The meaning of 'most': semantics, numerosity and psychology. *Mind and language 24*(5), 554–585.
Rubin, E. (1921). *Visuell wahrgenommene Figuren*. Copenhagen: Gyldendals.
Schlenker, P. (2008). Be articulate: a pragmatic theory of presupposition projection. *Theoretical linguistics 34*, 157–212.
Stalnaker, R. C. (1974). Pragmatic presuppositions. In M. K. Munitz and P. Unger (Eds.), *Semantics and philosophy*, pp. 197–214. New York University Press.
van der Sandt, R. (1992). Presupposition projection as anaphora resolution. *Journal of semantics 9*, 333–377.

Appendix: Further details on experimental design and data

Experiment 1

Participants

We posted surveys for 88 participants on Amazon's Mechanical Turk (mean age: 39; range: 17–77; 47 females). Only workers with an IP address from the United States were eligible for participation. These workers were asked to specify their native language, but payment was not contingent on their answer to this question. Two participants were excluded from the analysis because they were not native speakers of English.

Materials and procedure

A trial consisted of a sentence and a picture. Participants were instructed to indicate if the sentence was true or false as a description of the corresponding picture. Participants could also choose the answer "Don't know". The full instructions went as follows:

On each of the following pages of this survey, you will see a picture and a sentence. In each case, we ask you to decide whether the sentence gives a correct description of the picture or not. If it does, check "True". If it doesn't, check "False". If you feel you cannot decide whether the sentence is true or false, check "Don't know".

The survey consisted of 15 items: 5 targets and 10 fillers. The sentences in both conditions were of the form "Q {circle has/circles have} …". The form of the quantifier Q was varied between participants: plain (24 participants), D-partitive (24 participants), or DN-partitive (40 participants). The corresponding pictures consisted of four squares and five circles on a 3×3 grid. The shapes were coloured red, green, or blue.

The target sentences were of the following form:

Q {circle has/circles have} the same color as the square to which {it is/they are} connected.

In the corresponding pictures, four of the circles were each connected to one square, while the remaining circle was unconnected. The colours were distributed so that the sentence was false on a universal construal of the presupposition and true otherwise. For each of the quantifiers, five pictures were constructed, varying the distribution of the colours, the connections between circles and squares, and the position of the unconnected circle. Sample target pictures for each of the quantifiers are given in Figure 6.

Filler sentences were structurally similar to the target sentences. Three examples with plain quantifiers are:

(1) a. Every circle is connected to a circle of a different color.
    b. More than two squares are connected to less than two circles.
    c. Less than three squares are connected to a red circle.

The corresponding pictures for the filler items unequivocally verified or falsified these sentences.

Five lists were created, varying the order of the items and the correct responses to the filler items. The first two items were always fillers, and target items were always separated by at least one filler item.

Results

Filler items were answered correctly 79% of the time. One response was missing. The answer "Don't know" was extremely rare (< 1%). Since it is

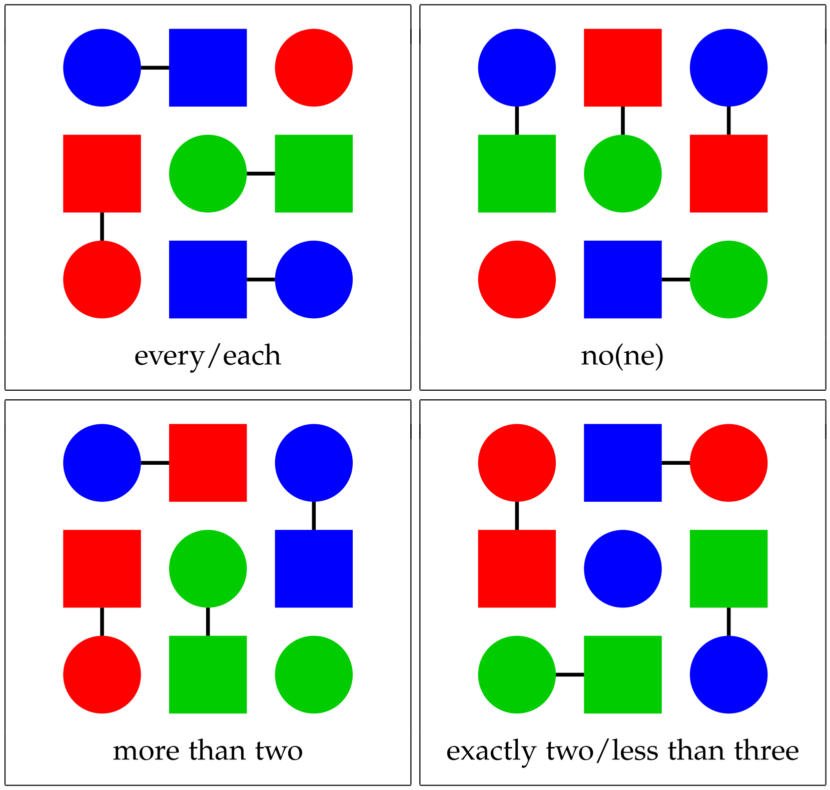


Figure 6: Example pictures used in Experiment 1.

unclear how to interpret these responses, they were discarded from the analysis. The percentages of "True" responses are given in Table 3.

	Plain Qs	D-partitive Qs	DN-partitive Qs
every/each	87	87	13
no(ne)	96	100	92
exactly two	91	91	90
more than two	87	87	98
less than three	100	100	89

Table 3: Percentages of participants in Experiment 1 who indicated that the target sentence was true in a situation falsifying the universal inference.

We compared the proportions of positive responses by means of pairwise Z-tests. None of the proportions of positive responses were significantly different (all p's $> .05$), except for the proportion of positive responses for "each of these n", which differed significantly from the proportions of positive responses for all other quantifiers (all p's $< .001$). Furthermore, there was a highly significant correlation between the number of incorrect answers to filler items and the number of "False" responses to target items ($r = .51$, $t(84) = 5.37$, $p < .001$). So the more errors participants made in the filler items, the more likely they were to give "False" responses to target items. This correlation suggests that some of the "False" responses to target items may be attributed to mistakes.

Experiment 2

Participants

We posted surveys for 50 participants on Amazon's Mechanical Turk (mean age: 30; range: 18–59; 17 females). Only workers with an IP address from the United States were eligible for participation. These workers were asked to specify their native language, but payment was not contingent on their answer to this question. All participants turned out to be native speakers of English.

Materials and procedure

A trial consisted of a sentence and a picture. Participants were instructed to indicate if the sentence was true or false as a description of the corresponding picture. The instructions went as follows:

> In the following survey, we will show you pairs of sentences and pictures. In each case, we ask you to decide whether or not the sentence gives a correct description of the picture. If you feel that the sentence is true, check "True". If not, check "False".
>
> We are interested in your spontaneous judgments, so please don't think too long about your answers.

The survey consisted of five items: 1 target and 4 fillers. The target sentence was "All of the squares are red". The corresponding picture alternated between picture A (25 participants) and picture B (25 participants) from Figure 3. Filler items were either ambiguous, involved some kind of visual illusion, or required analytic thinking. An example of the last category is shown in Figure 7. One list was created. The target item was the fourth item in the list.

Results

Participants were divided about the ambiguous and illusory filler items. All participants gave the correct answer to the filler item in Figure 7. Responses to the target item depended on the picture. For picture A, 52%

pink blue blue red

Only one of these words names the color of an adjacent word.

☐ True ☐ False

Figure 7: Example of a filler item used in Experiment 2.

of the participants judged that the sentence was true; for picture B, 16% did. This difference was statistically significant ($Z = 2.69$, $p < .01$).

Experiment 3

Participants

We posted surveys for 84 participants on Amazon's Mechanical Turk (mean age: 34; range: 20–80; 44 females). Only workers with an IP address from the United States were eligible for participation. These workers were asked to specify their native language, but payment was not contingent on the answer to this question. Three participants were excluded from the analysis because they were not native speakers of English.

Materials and procedure

A trial consisted of a sentence and a picture. Participants were instructed to indicate if the sentence was true or false as a description of the corresponding picture. Participants could also choose the answer "Don't know". The instructions were the same as for Experiment 1.

We constructed three types of pictures: one with seven circles and seven squares on a 2×7 grid and one with five circles and five squares on a 3×4 grid. There were two kinds of 3×4 pictures, depending on whether two (= 3×4|2) or three (= 3×4|3) colours were used. In the 3×4 pictures, there was one unconnected circle. We varied whether this circle had the same colour as the corresponding unconnected square, but this didn't have an effect. In the 2×7 pictures, five circles were unconnected. Some but not all of these had the same colour as the corresponding square.

Examples of the three types of pictures are shown in Figures 4 and 5. For each picture type, we constructed five pictures, varying the selection and distribution of the colours, and the position of the unconnected circle or circles. There were three groups of participants, one or each type of picture: 20 participants saw the 3×4|3 pictures, 24 participants saw the 3×4|2 pictures, and 40 participants saw the 2×7 pictures.

The target sentences were the same as in Experiment 1, but we only tested DN-partitive quantifiers. So all target sentences were of the following form, with *n* being the total number of circles in the picture:

> Q of these *n* circles {has/have} the same color as the square to which {it is/they are} connected.

For the $3{\times}4|3$ pictures, we tested all five DN-partitive quantifiers listed in Table 1, along with ten fillers. Since the pattern of results was similar to that of Experiment 1, we only tested the quantifiers "each of these *n*" and "none of these *n*" for the other two picture types, reducing the number of fillers to eight. In the surveys with $3{\times}4|2$ and $2{\times}7$ pictures, we also included two control items to gauge if participants correctly parsed the target sentences and pictures. These were of the following form, and were paired with pictures that made the sentences unambiguously true or false:

> Q of these *n* circles which is connected to a square has {the same/a different} color {as/than} the square to which it is connected.

We created four lists of $2{\times}7$ items, five lists of $3{\times}4|3$ items, and eight lists of $3{\times}4|2$ items, varying the order of the items and the correct responses to the filler items. The first two items were always fillers, and target items were separated by at least one filler item.

Results

Filler items were answered correctly 79% of the time. Control items were answered correctly 92% of the time. The answer "Don't know" was extremely rare ($< 4\%$). Since it is unclear how to interpret these responses, they were discarded from the analysis. The results for the quantifiers "each of these *n*" and "none of these *n*" are shown in Table 2.

The distribution of answers for "none of these *n*" was similar to the results for that quantifier in Experiment 1. The distribution of answers for "each of these *n*" differed depending on the type of picture. The proportion of positive responses for the $3{\times}4|3$ pictures (35%) was significantly higher than for the $3{\times}3$ pictures tested in Experiment 1 (13%, $Z = -1.99$, $p = .047$), and significantly lower than for the $3{\times}4|2$ pictures (68%, $Z = 2.05$, $p = .040$) and the $2{\times}7$ pictures (65%, $Z = 2.07$, $p = .039$). The difference between the proportions of positive responses for the last two pictures was not statistically significant ($Z < 1$).