

Embedded scalars*

Bart Geurts
University of Nijmegen

Bob van Tiel
University of Nijmegen

Submitted 2012-06-26 / Accepted 2012-08-07 / Revisions received 2012-12-26 /
Corrections requested 2013-01-19 / Final version received 2013-01-24 / Published
2013-10-01

Abstract For over a decade, the interpretation of scalar expressions under embedding has been a much debated issue, with proposed accounts ranging from strictly pragmatic, on one end of the spectrum, to lexico-syntactic, on the other. There has been some confusion as to what exactly the controversy is about, and we argue that what is at stake is the division of labour between pragmatic and truth-conditional mechanisms. All parties to the debate agree that upper-bounded construals of scalar expressions are variously caused by conversational implicatures and truth-conditional narrowing, but whereas Griceans argue that the former mechanism is the main cause, conventionalists point to the latter, assuming as a matter of course that the source of truth-conditional narrowing lies in linguistic convention; on this view, narrowing is either a lexical or a syntactic phenomenon. Since researchers' introspective judgments tend to agree with the theories they advocate, a number of experimental studies have recently tried to shed light on this issue. In this paper, we review the experimental record, and argue that the extant data favour a pragmatic account.

Keywords: scalar implicatures, quantification, typicality

* We are indebted to Emmanuel Chemla and Benjamin Spector for giving us permission to use their raw experimental data, and to Corien Bary, Chris Cummins, Larry Horn, Louise McNally, Sammie Tarenskeen, Rob van der Sandt, Noor van Leusen, Emiel van Miltenburg, and an anonymous reviewer for *S&P*, for their extremely useful comments and questions. This research was supported by a grant from the Netherlands Organisation for Scientific Research (NWO).

Introduction

Well before “Logic and conversation” went into print (Grice 1975), Cohen (1971) pointed out a problem with Grice’s account of conversational implicature: occasionally, what would appear to be a conversational implicature is observed in an embedded position — which is a contradiction in terms, since conversational implicatures can only be derived on the basis of a full-blown speech act. The debate about this issue continues until the present day, and in recent years has homed in on the interpretation of scalar expressions under embedding. To explain what is at stake, consider the following example:

- (1) Julius believes that some of the goats are happy.

An utterance of this sentence may give rise to the inference that, according to the speaker, Julius believes that not all the goats are happy. This inference involves two layers of belief:

- (2) $BEL_S(BEL_J(\neg(\text{all the goats are happy})))$

The problem is that, if we treat this as a regular case of scalar implicature, the outcome will be somewhat weaker:

- (3) $BEL_S(\neg BEL_J(\text{all the goats are happy}))$

(Note that (2) entails (3), but not vice versa.) Here is how the inference in (3) comes about. Upon hearing (1), the addressee reasons as follows:

- i. Could it be the case that, for all the speaker knows, Julius believes that all the goats are happy, i.e. $BEL_S(BEL_J(\text{all the goats are happy}))$?
- ii. Presumably not, for then the speaker would have said, “Julius believes that all the goats are happy.” Hence, $\neg BEL_S(BEL_J(\text{all the goats are happy}))$.
- iii. It seems reasonable to suppose that the speaker is competent with respect to the proposition that $BEL_J(\text{all the goats are happy})$. That is, either $BEL_S(BEL_J(\text{all the goats are happy}))$ or $BEL_S(\neg BEL_J(\text{all the goats are happy}))$.
- iv. If so, it follows that $BEL_S(\neg BEL_J(\text{all the goats are happy})) (= (3))$.

This result is not wrong, and may be quite satisfactory in most cases, but it falls short of capturing the inference in (2).

This particular problem was noted by Landman (1998), who used it to argue for a wholesale departure from the Gricean party line. Landman’s cause was joined by Chierchia (2004), who threw in some problem cases of his own devising, with scalar expressions occurring in the scope of disjunctions, factive verbs, and existential indefinites. The challenge posed by Landman and Chierchia provoked a flurry of Gricean responses by Sauerland (2004), van Rooij & Schulz (2004), Horn (2006, 2009), Spector (2006), Russell (2006), and Geurts (2009, 2010). This round in the debate showed that, *prima facie* impressions notwithstanding, Landman’s and Chierchia’s alleged counterexamples can be accounted for within a Gricean framework, after all. To illustrate, in a context in which (1) is uttered, it is likely to be common ground that either Julius believes that all the goats are happy or believes that not all the goats are happy; or more succinctly:

- (4) $BEL_j(\text{all the goats are happy}) \vee BEL_j(\neg(\text{all the goats are happy}))$

Taken together with the implicature in (3), this assumption entails (2) (van Rooij & Schulz 2004, Russell 2006, Geurts 2009, 2010). On this analysis, Landman’s reading of (1) doesn’t require a local upper bound on the interpretation of *some*. Rather, it follows from an ordinary quantity implicature provided (4) is part of the common ground.

The Gricean strategy for dealing with Landman and Chierchia’s observations is to proceed case by case: while all their examples can be treated in a principled manner (as we just illustrated for belief sentences), there is no common cure. Put otherwise, on the Gricean account, all “embedded implicatures” are special in some way or other; they are rare and there is no uniform explanation for their occurrence in the scope of *believe*, *or*, and so on. This is an important point, for the central tenet of the critics of the Gricean programme is that upper-bounded construals of scalar expressions occur across the board, be they embedded or not. Put otherwise, anti-Griceans see a continuity between embedded and non-embedded contexts that Griceans cannot accept.

Before we continue our story, there are some terminological issues to clear out of the way. In the recent literature, “implicature” has been watered down to a purely descriptive term. (Much to his regret, the first author has been guilty of this deplorable practice, too.) We believe that this usage has been a source of much confusion, and will therefore avoid it here. Implicatures are conversational inferences, and it is pointless even to consider the possibility that an implicature might occur in an embedded position. Instead we will

speak of “upper-bounded construals” of scalar expressions, or UBCs for short, whenever we wish to remain neutral on the underlying mechanisms. A UBC may be either an implicature or a truth-conditional effect, and it is only in the latter case that it can occur in the scope of, e.g., a conditional, a quantifier, or negation.

1 What the debate is *not* about

With the first volley of problem cases cleared out of the way, the anti-Griceans changed tack, and shifted their focus to examples like the following (all from Chierchia, Fox & Spector 2012):

- (5) a. If you take salad OR dessert, you pay \$20; but if you take BOTH there is a surcharge.
 b. Exactly three students did MOST of the exercises; the rest did them ALL.
 c. It is not just that you CAN write a reply. You MUST.

Oddly enough, however, advocates of the Gricean programme have been perfectly happy to concede that *these* data cannot be accounted for in terms of conversational implicature (e.g., Horn 1989, 2006, 2009, Geurts 2009, 2010, Geurts & Pouscoulous 2009). In fact, none of the data mustered by Chierchia, Fox & Spector (2012) in their broadside attack on Gricean pragmatics are particularly worrisome for the defenders of that view. So what, if anything, is the current debate about?

In the debate between Griceans and their critics, there has been an unfortunate misunderstanding on the part of the latter. The misunderstanding is that, for those of us who prefer to take the Gricean approach, there is no way to account for the data in (5). Not true. What *is* true is that there is no way of explaining these data in terms of honest-to-Grice conversational implicatures. But that just means they must be something else. It is part and parcel of the Gricean view that there are two routes for a scalar expression to arrive at a UBC. One is by way of a quantity implicature; the other, by way of truth-conditional narrowing (which, as we will presently see, can be construed in pragmatic terms, as well). So in (5c) *can* means “can but not must”, and the same, *mutatis mutandis*, for (5a) and (5b). Crucially, such embedded UBCs are always marked, as the examples in (5) illustrate.

Hence, according to the Gricean view, there are two mechanisms that variously underwrite UBCs: implicature and truth-conditional narrowing. This

is one point the critics of the Gricean programme have glossed over. Another is that *this assumption is common ground between the parties in the debate*. That's right: there is a broad consensus that the dual-mechanism view is correct.¹ Fox (2007) is exceptionally lucid on this point, but even Chierchia, Fox & Spector (2012), who mount a hard-nosed defence of truth-conditional UBCs, admit in a footnote that in some cases quantity implicatures must be appealed to (Chierchia, Fox & Spector 2012: note 35). Everybody agrees, for example, that in some contexts, an utterance of (6) will merely imply that the speaker doesn't know whether Barney took all the strawberries:

(6) Barney took some of the strawberries.

This inference, which is weaker than the usual $BEL_S(\neg(\text{Barney took all the strawberries}))$, cannot be explained as a case of truth-conditional narrowing by any theory, and therefore has to be a quantity implicature, as Chierchia et al. concede.

2 What the debate *is* about

Given that there is a broad consensus about the mechanisms underlying UBCs of scalar expressions, what is the controversy about? As discussed at length by Geurts (2010), there are two main issues. One concerns the aetiology of truth-conditional narrowing: what is its underlying cause? This question has given rise to a variety of answers, running a gamut of potential sources ranging from the lexicon via syntax to pragmatics. However, thus far there has been hardly any discussion of the relative merits of the various approaches. Instead, the debate has focused on the second question, which concerns the division of labour between implicature and truth-conditional narrowing. According to the critics of the Gricean approach, truth-conditional narrowing applies across the board, and therefore we should expect to observe UBCs all over the place, regardless of whether scalar expressions are embedded or not. By contrast, on the Gricean view, while scalar implicatures cannot be embedded by their very nature, truth-conditional narrowing is a marked option, and therefore embedded UBCs are predicted to be exceptional. As emphasised by Ippolito (2010), this has been the main issue in the literature

¹ See, e.g., Bach (1994), Recanati (2003), Horn (2006, 2009), Fox (2007), Noveck & Sperber (2007), Geurts (2009, 2010), Ippolito (2010), Chierchia, Fox & Spector (2012).

on so-called “embedded implicatures”, and it is what the bulk of this paper is about.

In order to get a proper perspective on the ensuing discussion of experimental evidence, which is our main concern, this section elaborates on both issues, though it may be useful to keep in mind that our chief topic is the second one, viz. the division of labour between implicature and truth-conditional narrowing.

2.1 What underlies truth-conditional narrowing?

This question has been answered in three markedly distinct ways, two of which are quite similar in spirit in that they share the presupposition that narrowing is a linguistic phenomenon; we therefore lump them together under the “conventionalist” label. The third view is pragmatic.

- CONVENTIONALIST
 - *Lexicalist*
Levinson (2000), Chierchia (2004), Storto & Tanenhaus (2005)
 - *Grammatical*
Chierchia (2006), Fox (2007), Chierchia, Fox & Spector (2012)
- PRAGMATIC
Horn (2006, 2009), Noveck & Sperber (2007), Geurts (2009, 2010)

Technically as well as conceptually, the lexicalist approach is the simplest, the idea being that scalar expressions come with an upper-bounded meaning by default: *some* just means “some but not all”, where the “not all” part is cancellable. The grammatical approach to narrowing is rather more convoluted, and of a more recent vintage. It is based on the idea that a scalar expression may acquire a UBC by sitting in the scope of a covert syntactic operator whose meaning is essentially that of overt *only*. When in the scope of this covert operator, *some* will be interpreted, in effect, as “only some”, and thus come to exclude “all”.

The pragmatic view on truth-conditional narrowing is quite different. It is based on the observation that, in general, the occurrent meaning of a word does not coincide with its dictionary meaning, but is modulated by contextual factors.² This process may produce all sorts of meanings,

² See, e.g., Nunberg (1978, 1979), Sag (1981), Bosch (1984), Bartsch (1987), Recanati (1986, 2004), Geurts (2009, 2010).

and occasionally the occurrent meaning of a word will be more specific than its dictionary meaning. In some cases, this narrowing may proceed quite smoothly, because world knowledge alone suffices to steer the hearer towards a specific interpretation. In many other cases, however, narrowing must be helped along by contrastive stress, as in the following examples from Geurts (2010: 186):

- (7) a. When he DRINKS, he DRINKS.
- b. Julius isn't RICH: he's RICH.
- c. Cleo didn't OPEN her handbag: she made an INCISION on the side.

None of these examples can plausibly be accounted for in conventionalist terms, for in each case, the occurrent meanings of the key expressions are heavily dependent on the context. According to the pragmatic view, this holds for truth-conditional narrowing across the board, including the examples in (5), which are the mainstay of the recent conventionalist critique of the pragmatic approach to UBCs.

Though the bandwidth of theories of narrowing is considerable, as we have just seen, and it is important to realise that there is more than one game in town, in the following pages we will confine our attention to the conventionalist account, since this is the one advocated by critics of the Gricean approach to scalars: as we will presently see, the key issue in the experimental literature on embedded UBCs is whether or not truth-conditional narrowing is a freely available option in any syntactic position, as conventionalist accounts predict.

2.2 The division of labour between narrowing and implicature

Let's develop the main positions on embedded UBCs a bit further by focusing on an example over which opinions are strongly divided (Geurts & Pouscoulous 2009):

- (8) All the squares are connected with some of the circles.

An utterance of (8) may give rise to the following inference:

- (9) Not all the squares are connected with all of the circles.

This much is agreed upon, and it is also agreed that (8) can give rise to a reading on which it is truth-conditionally equivalent to:

- (10) All the squares are connected with some but not all of the circles (and therefore none of the squares are connected with all of the circles).

However, the Gricean account predicts that this reading is available only marginally, and needs to be forced, e.g., by contrastive stress on *some* in (8).

By contrast, according to the conventionalist doctrine, UBCs “occur systematically and freely in arbitrarily embedded positions.”³ On this view, in the context of a sentence like (8), *some* will receive a truth-conditional UBC by default, and hence this sentence will normally be read as being equivalent with (10).

This default is accounted for in different ways, depending in part on how the question about the origin of UBCs is answered. If *some* is held to be lexically ambiguous between “at least some” and “at least some but not all”, it may be stipulated that the latter reading is the default, or alternatively, it may be stipulated that UBCs are preferred under such-and-such circumstances, e.g., when a scalar expression doesn’t occur in a downward-entailing environment (Landman 1998, 2000, Chierchia 2004). If, on the other hand, a grammar-based version of conventionalism is adopted, another approach is called for. Chierchia (2006) suggests that whenever *only*-insertion leaves us with several parses to choose from, the one with the strongest meaning is preferred, and various ways of fleshing out this idea are canvassed by Chierchia, Fox & Spector (2012). In the case of (8) this theory, too, predicts that (10) is the preferred reading. More generally, conventionalist theories agree that, at least in upward-entailing environments, there is a preference for interpreting embedded *some*, in effect, as “some but not all”, even if they disagree about how the construal and the preference come about.

According to our intuitions, none of the following inferences would go through under normal circumstances:

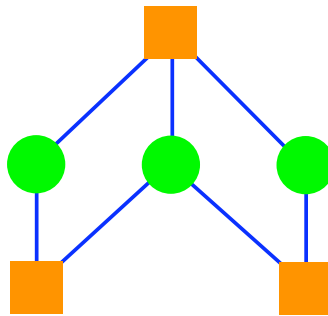
- (11) a. All the villagers own rabbits or chickens.
 ↯ None of them own both.
 b. At least 200 of the villagers own rabbits or chickens.
 ↯ At least 200 of them don’t own both.

³ The quote is from Chierchia, Fox & Spector (2012), who go so far as claiming that this is an “empirical generalization”. There is an obvious tension between Chierchia et al.’s statement that embedded UBCs occur “systematically and freely” and the fact that all the examples they cite in its support are marked and require contrastive stress. Chierchia et al. seem to be unaware of this tension. In fact, as noted by Sauerland (2010), Chierchia et al.’s views are contradictory in other ways, too.

- c. Only six of the villagers own rabbits or chickens.
 \nrightarrow Only six of the villagers own rabbits or chickens but not both.

However, conventionalist theories generally predict that (11a) and (11b) should hold by default, and at least some of them predict this for (11c) as well. In (11a,b), *or* occurs in an upward-entailing environment, and therefore most conventionalist theories concur that an embedded UBC of the disjunction should be preferred; i.e. *or* should be read exclusively, as meaning “or but not and”. In (11c), *or* occurs in a non-monotone environment (i.e., it is neither upward nor downward entailing), and in such cases a bias towards an exclusive reading of *or* is predicted by at least some variants of conventionalism.

In the remainder of this paper, we will discuss experimental evidence for and against the predictions that UBCs of scalar expressions are freely available in (a) upward-entailing contexts (§§3-5) and (b) non-monotone contexts (§6). We will see that neither prediction is borne out by the data.



All the squares are connected
with some of the circles.

true false

Figure 1: Experimental item used by Geurts & Pouscoulous (2009).

3 Experimental evidence against conventionalism

The first experimental study to test the aforementioned predictions was by Geurts & Pouscoulous (2009), who did a verification experiment with sentences like the following:

- (12) a. All the squares are connected with some of the circles. (= (8))
 b. There is more than one square that is connected with some of the circles.
 c. There are exactly two squares that are connected with some of the circles.

Each of these sentences was paired with a picture for which the sentence's standard and conventionalist construals yielded conflicting truth values. For example, when interpreted according to the conventionalist canon, the sentence in Figure 1, i.e. (12a), fails to match the depicted situation, because one of the squares is connected to all the circles, but the sentence is true on its standard reading.

G&P's findings were as univocal as they get: not a single one of their participants' responses agreed with conventionalist predictions. For example, participants invariably gave a positive response in the case of (12a) (Figure 1), and the same held, *mutatis mutandis*, for the other conditions, regardless of whether they were upward entailing or merely non-downward-entailing.

These findings robustly fail to confirm what G&P call "mainstream conventionalism": a range of theories which predict that scalar expressions preferably receive an upper-bounded construal in upward-entailing environments and perhaps in non-downward-entailing environments generally. Since these predictions weren't borne out by their results, G&P went on to test a weaker version of conventionalism, which refuses to deal with the ambiguities it begets, and doesn't make any predictions about which construal is the preferred one. According to this minimal variety of conventionalism, each of the sentences in (12) has several readings on any given occasion, but it is for pragmatics to decide which is the right one.⁴

Obviously, in its most austere form, minimal conventionalism fails to make any predictions at all: the claim that, e.g., (13a) may or may not be read as (13b) cannot be falsified.

- (13) a. All the squares are connected with some of the circles.
 b. All the squares are connected with some but not all of the circles.

⁴ To the best of our knowledge, nobody has as yet come forward to defend minimal conventionalism. Chemla and Spector disagree, claiming as they do that this view "can be attributed to" Chierchia (2006), Fox (2007), Magri (2009), and Chierchia, Fox & Spector (2012) (Chemla & Spector 2011: 363; cf. Sauerland 2010). However, Chemla and Spector's "can" can only be read counterfactually: none of the authors they mention explicitly align themselves with the minimalist view, and all of them at least imply that they don't.

In order for the theory to have at least a modicum of empirical bite, auxiliary assumptions will have to be made. One hypothesis that naturally comes to mind is that speakers are able to *detect* the readings that are characteristic of the conventionalist approach. That is to say, even if speakers don't prefer to hear (13a) as synonymous with (13b), it is reasonable to expect them to realise that the latter sentence expresses a possible construal of the former. If this auxiliary hypothesis is adopted, minimal conventionalism predicts that, in a situation that falsifies (13b), native speakers should judge that (13a) is ambiguous between a true reading and a false one. This prediction was tested in a further experiment by [Geurts & Pouscoulous \(2009\)](#).

For this experiment, G&P used the same materials as in the verification study discussed above, but added a third response option: participants were invited to choose between “true”, “false”, and “could be either”. According to minimal conventionalism, participants should have a preference for the last option in the upward-entailing conditions, and perhaps in non-downward-entailing conditions generally. In order to assess whether participants can cope with this type of task in the first place, G&P included a number of control sentences which were genuinely ambiguous; an example is (14), which can be read as (14a) or (14b):

- (14) There are green circles and squares.
- a. There are green circles and there are squares.
 - b. There are green circles and there are green squares.

The main finding of this study was that, whereas ambiguous controls were recognised as such in 70% of the cases, the corresponding rate for the non-downward-entailing sentences was 6%, while for the upward-entailing sentences it was 0%. On a further 5% of the non-downward-entailing trials, participants gave a true/false response that was consistent with a conventionalist construal. In raw numbers, only 9 out of 88 responses were consistent with minimal conventionalism. Moreover, all but one of these “positive” (or better: non-negative) responses were associated with *exactly two*: between them, upward-entailing *all* and *more than one* prompted a conventionalist response on *one* trial only.

In short, while they were rather good at picking out undisputedly ambiguous sentences, participants overwhelmingly failed to produce responses consistent with a conventionalist analysis of scalar expressions, either by classifying target sentences as ambiguous or by assigning them a truth value

in line with a conventionalist construal. In our opinion, this is *very* strong evidence against the continuity hypothesis endorsed by conventionalists: embedded and unembedded scalars are interpreted quite differently.

4 Conventionalist objections

Geurts and Pouscoulous’s experiments have been discussed and criticised by Clifton & Dube (2010), Ippolito (2010), Sauerland (2010), and Chemla & Spector (2011). It would take us too far afield to consider all the issues raised by these authors, so in the following we will confine our attention to what we take to be the main bones of contention. To begin with, we consider Chemla & Spector’s (2011) complaint that the pictures used in Geurts and Pouscoulous’s experiments were “difficult to decipher”:

Consider the example depicted in Figure 1 again. The crucial bit of information for the present purposes is that the square on top of the picture is connected with all the circles, hereby falsifying the [embedded UBC]. On purely introspective grounds, we find this information pretty hard to extract, and participants may either miss it or ignore it altogether. (Chemla & Spector 2011: 365)

This is a curious objection. For one thing, the pictures used by Chemla and Spector in their own experiment (cf. Figure 4 below) were, if anything, more complex than Geurts and Pouscoulous’s. For another, it is easily shown that the participants in Geurts and Pouscoulous’s experiments had no problems parsing pictures like the one in Figure 1. First, had there been problems in this respect, they should have manifested themselves also in the filler and control items used by Geurts and Pouscoulous. Nothing in their data suggests that this may have been the case. Secondly, on the conventionalist view, the preferred reading of (15a) is (15b):

- (15) a. All the squares are connected with some of the circles.
 b. All the squares are connected with some but not all of the circles.

It bears emphasising, perhaps, that on the view under discussion, the preferred *truth conditions* of the (a)-sentence are the same as those of the (b)-sentence. Now, is (15b) true or false in the situation depicted in Figure 1? We believe it is obviously false, but in order to verify our intuitions, we asked

27 students the same question in a classroom setting. 25 of our informants agreed, without any hesitation, that the sentence is false. Since none of the participants in Geurts and Pouscoulous's experiments judged (15a) to be false, in the same situation, there is a stark contrast here. Apparently, pictorial complexity was not an issue in these studies.

Another objection raised by Chemla and Spector has to do with relevance:

We speculate that another reason why the local reading might have been particularly hard to detect in Geurts and Pouscoulous' task, even if it existed, is that the pictures they used failed to make the local reading sufficiently *relevant*. (Chemla & Spector 2011: 365)

We fail to see why the pictures in a verification task should be held responsible for selecting a suitable reading, but that is as it may be, for in the present context, relevance is a red herring. In the first experiment discussed above, Geurts and Pouscoulous tested the claim that embedded scalars receive a UBC *by default*; in the second experiment, the hypothesis under consideration was that, even if they are not preferred, embedded UBCs are sufficiently salient to be detected *under neutral conditions*. In either case, relevance is quite simply irrelevant.

The third and last objection we would like to consider takes its inspiration from the so-called "Principle of Charity", which was first formulated by Wilson (1959), and popularised by Quine (1960) and Davidson (1973). In the philosophical literature, the Charity Principle has taken on many different forms, but for our purposes, the most relevant one is the following, which is the hearer's counterpart to Grice's (1975) Quality Maxim:

(16) *Principle of Charity*

Try to interpret the speaker's utterance in such a way that it is true.

One way of trying to save the conventionalist account in the face of Geurts and Pouscoulous's data is by arguing that something akin to the Principle of Charity masks the unattested reading according to which (15a) conveys (17b) rather than merely the weaker (17a):

- (17) a. $\forall x[Sx \rightarrow \exists y[Cy \wedge Rxy]]$ (literal meaning)
b. $\forall x[Sx \rightarrow (\exists y[Cy \wedge Rxy] \wedge \exists y[Cy \wedge \neg Rxy])]$ (with embedded UBC)

(Chemla & Spector 2011: 366) suggest that “in some cases where a sentence is ambiguous between two readings R1 and R2, where R2 asymmetrically entails R1, naive subjects are not even aware of an ambiguity.” However, even if this claim were true, it would not be sufficient to block the reading in (17b). If we want to argue that the availability of the reading in (17a) suppresses (17b) as a candidate interpretation, something like the following principle should hold with sufficient generality:

(18) *Preference for Truth*

If a sentence is ambiguous between two readings R1 and R2, where R2 asymmetrically entails R1, then naive subjects will only perceive reading R1.

“Preference for Truth” is the name Chemla and Spector use to refer to (or rather hint at) a Charity-like principle which they don’t define. (18) is an attempt to capture what they need. If this principle could be motivated on independent grounds, then it might explain why sentences like (15a) don’t seem to license embedded UBCs. However, unlike the Charity Principle, its false friend in (18) is not plausible at all; for it is glaringly false to suppose that, in general, stronger readings are rendered inaccessible by the availability of weaker competitors. This fact is nicely illustrated by some of the control sentences in Geurts and Pouscoulous’s ambiguity experiment, like (14), for example, which was recognised as ambiguous by 77% of Geurts and Pouscoulous’s participants;⁵ more counterexamples are easy enough to find.⁶ Furthermore, it is just as false to suppose that, in the case of scalar expressions, UBCs are systematically suppressed by weaker readings. Chemla and Spector themselves expressly deny that this is so, even going so far as to suggest that “there is an overall preference for deriving [UBCs], independently of considerations of logical strength.” (p. 366) Though this is a pretty wild exaggeration, as the experimental record fails to support a general bias

⁵ Surprisingly, Chemla and Spector claim (pp. 366-367), contrary to fact, that none of Geurts and Pouscoulous’s control sentences were ambiguous between stronger and weaker construals.

⁶ In his review of our paper for *S&P*, Larry Horn mentions the following cases in which a stronger reading is preferred over a weaker one, or there is at least a palpable ambiguity between the two:

- Quantifier scoping: “Everyone didn’t solve the problem.”
- Neg-raising: “I don’t think they’ll survive.”
- Lexical: “He’s drinking again.” [alcohol or liquids]

For more examples and further discussion of this theme, see Geurts (2000).

for UBCs (see [Geurts 2010](#) for extensive discussion), there can be no doubt that, in unembedded contexts at least, scalar expressions give rise to UBCs often enough, which proves that they are not systematically suppressed by the availability of weaker readings. In short, (18) is untenable as a general principle and yields obviously incorrect predictions for quantity implicatures, in particular. So this turns out to be a cul-de-sac.

[Sauerland \(2010\)](#) suggests another way in which something similar to the Principle of Charity might serve to reconcile conventionalism with the facts. He appeals to a principle proposed by [Meyer & Sauerland \(2009\)](#), which goes as follows:

(19) *Truth Dominance*

Whenever an ambiguous sentence *S* is true in a situation on its most accessible reading, we must judge sentence *S* to be true in that situation. ([Meyer & Sauerland 2009: 140](#))

The accessibility rider is of course crucial. What is the “most accessible reading” of our running example?

(20) All the squares are connected with some of the circles.

According to the particular version of conventionalism favoured by Sauerland, the strongest interpretation of (20) involves a covert *only*-operator whose focus is *some*; on this construal, the sentence is read, in effect, as (17b). Sauerland hypothesises that this reading is less accessible than the literal meaning of (20) precisely because it involves an additional (covert) operator, and thus Truth Dominance blocks the reading in (17b).

Sauerland’s account runs into the same problems as Chemla and Spector’s. First, since in Sauerland’s version of conventionalism, regular quantity implicatures always require more silent operators than literal meanings, his analysis entails that quantity implicatures should be systematically suppressed even in the simplest cases:

(21) Some of the nurses are snoring.

↷ Not all the nurses are snoring.

If this prediction were correct, the entire literature on UBCs of scalar expressions would be on the wrong track. Secondly, pace Sauerland, the independent

evidence for Truth Dominance is tenuous at best.⁷ On the contrary, it is clear that, if it is not circular to begin with, then like Chemla and Spector’s Preference for Truth, the principle proposed by Meyer and Sauerland rules out of court all manner of ambiguities that are readily observable (see note 6).

The Principle of Charity enjoins the hearer to try and interpret the speaker’s utterances in such a way that they are true. It is entirely plausible that, in general, hearers abide by this principle, but there is nothing in it to suggest that hearers will fail to detect an interpretation if there is another reading that is weaker or more accessible (whatever that may mean). The principles postulated by Chemla and Spector and Meyer and Sauerland are ad hoc as well as false, and therefore ill-suited for explaining away the experimental evidence presented by Geurts and Pouscoulous.

5 The conventionalist counterevidence and its problems

In the foregoing, we examined various attempts at reinterpreting Geurts and Pouscoulous’s findings so as to bring them in line with conventionalist predictions. These attempts were found wanting. In this section, we turn

⁷ Meyer & Sauerland (2009) postulate this principle in order to explain why the German sentence in (i) can only be interpreted as (ia), while the reading in (ib) goes undetected, even though many accounts of scope taking predict that it exists:

- (i) Nur Maria liebt jeder.
 only Maria_{ACC} loves everybody_{NOM}
- a. $\forall y[y = m \leftrightarrow \forall xLxy]$ (only > every)
 b. $\forall x\forall y[y = m \leftrightarrow Lxy]$ (every > only)

Meyer and Sauerland assume, plausibly enough, that readings which mirror surface scope are more accessible than readings that don’t. Therefore, (ia) is the most accessible interpretation of (i), and since (ib) is logically stronger than (ia), Truth Dominance predicts that it must be suppressed.

But now consider the following variation on (i):

- (ii) Nur Maria liebt niemand.
 only Maria_{ACC} loves nobody_{NOM}

Like (i), (ii) admits of one construal only, and in this case, too, it is the one that mirrors surface scope: “Maria is the only one who is loved by nobody.” However, this reading is not entailed by the one in which *nobody* takes scope over *only*, and therefore Truth Dominance doesn’t explain why the latter is not detected. This is odd, for the data would seem to be exactly parallel, and if Truth Dominance cannot explain why (ii) admits of one reading only, why suppose that it constrains the interpretation of (i)?

to two studies which claim to provide experimental evidence in favour of a conventionalist treatment of UBCs: one by Clifton & Dube (2010), the other by Chemla & Spector (2011). To set the stage for the upcoming discussion, let us briefly recap what is at issue.

As explained in §2, the debate between Gricean and non-Gricean theories is not about the question of whether embedded UBCs exist or not; it is agreed that they do. The real issue is whether or not there are intrinsic differences between embedded and non-embedded occurrences of scalar expressions. In a Gricean framework, there is a profound difference between (22a) and (22b): while the first sentence can be interpreted, in effect, as expressing that some but not all of the girls were sick, it cannot be thus interpreted when in the scope of a quantifier, as in (22b), unless the speaker uses marked means to force this interpretation, as in (22c):

- (22) a. Some of the girls were sick.
b. Every morning, some of the girls were sick.
c. Every morning, *SOME* of the girls were sick.

On a conventionalist view, by contrast, there is a single mechanism that generates UBCs in all these sentences. This is not to say that, on such a view, we should expect to find the same rates of UBCs for each of these sentences, but it would be most inconvenient if we observed elevated rates of UBCs for (22a) and no embedded UBCs at all for (22b).⁸ Since this is exactly what Geurts and Pouscoulous found, their data are a major embarrassment for orthodox conventionalists.

As discussed in §1, cases like (22c) do not separate the Gricean sheep from the conventionalist goats, since in these cases a contrastive interpretation is forced, and the two bovid species agree that these may be instances of truth-conditional narrowing. This is an important point, because we will argue that in the experiments to be discussed in the following, contrast was always a key factor. Hence, our first point will be that these experiments have no bearing on the debate between Griceans and conventionalists. Furthermore, and this will be our second point, for the most part these experiments fail to provide evidence for embedded UBCs of any kind, be they marked or not. Rather, what they show is that scalar expressions give rise to typicality effects.

⁸ Unless, that is, there are independent reasons for assuming that UBCs are suppressed in certain embedded contexts. But as saw in the last section, such reasons still remain to be found.

5.1 The best shape (Clifton & Dube 2010)

In their experiment, Clifton & Dube (2010) used a multiple-choice task to probe the interpretation of scalar items: instead of asking for truth-value judgments, as in Geurts and Pouscoulous’s study, participants were presented with four situations and invited to select the one that was best described by a given sentence. The critical items came in two versions, one of which is illustrated in Figure 2. In this version, participants were given a choice between the following response options:

- i. a picture that verified the literal meaning of the target sentence (without an embedded UBC);
- ii. a picture that verified the literal meaning of the sentence as well as an embedded UBC;
- iii. “both”;
- iv. “neither”.

Please indicate which shape is best described by the sentence below:

All the squares are connected to some of the circles.

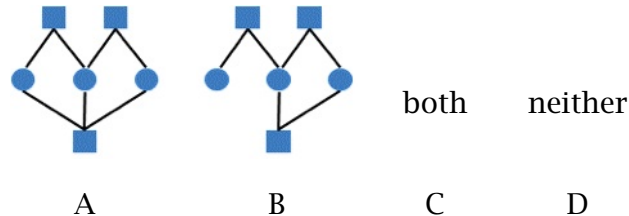


Figure 2: Experimental item used by Clifton & Dube (2010).

The second version was similar to the first, except that neither picture verified the embedded UBC. Instead, one picture was identical to the A-picture in Figure 2, while in the other picture, every square was connected to every circle.⁹ Hence, there were three different pictures altogether: in each case,

⁹ Clifton and Dube also varied the universal quantifier between *all* and *each*. Here, we only discuss the results for *all*; the data for *each* were not significantly different.

every square was connected to at least some of the circles, and in addition one of the following applied:

- P1: No square was connected to all circles.
- P2: Some though not all squares were connected to all circles.
- P3: Every square was connected to every circle.

<i>Response options</i>	P1	P2	P3	“both”	“neither”
Version 1	39	3	<i>n.a.</i>	57	1
Version 2	<i>n.a.</i>	28	6	50	17

Table 1: Percentages of choices for each option in Clifton & Dube’s (2010) experiment. Data points claimed as evidence for embedded UBCs are shaded.

The results are shown in Table 1. In both versions of the critical trials, the third option, “both”, was chosen in the majority of cases. In the first version, the P1-picture (= B in Figure 2) was a good second, and the remaining options were almost completely ignored. In the second version, the P2-picture (= A in Figure 2) came in second, but in this version there was a substantial minority in favour of “neither”. As indicated in Table 1, C&D assume, without argument, that these results prove that their participants preferred an embedded UBC in a significant number of cases. They claim that this interpretation prevailed in 39% and 17% of the trials, respectively. If this conclusion were justified, the experiment wouldn’t support the view that embedded UBCs are preferred, but at least it would show that they happen.

C&D’s experimental design raises a number of issues. First, there was a marked asymmetry between the pictorial A- and B-options, on the one hand, and the verbal C- and D-options, on the other. This asymmetry may have had an effect on the participants’ responses. For example, since the pictures were more salient than the verbal options, it is possible that there was a bias towards the A/B-options.

Secondly, it is unclear how C&D’s task had to be interpreted. Participants were asked to “indicate which shape is best described by” the target sentence. This phrasing suggests that one of the pictures is the best candidate. If so, how are the verbal options to be understood? “Both” could then be construed as “Both pictures are equally good”, which itself can mean different things,

depending on whether it is taken to imply that both pictures are good; or it could be construed as “Don’t know”. *Mutatis mutandis*, the same holds for the “neither” response.

Thirdly, C&D’s task presupposes that the target sentence has a preferred interpretation; for if it hadn’t, it is not clear what participants should do. Consider the trial in Figure 2 again and imagine a participant who manages to construe the target sentence both ways and has no distinct preference for either. How is she to decide between various options? By flipping a coin? In sum, C&D’s experimental design was not of the cleanest, and hence their findings are equivocal, at best.

Apart from these methodological reservations, it is quite evident that contrast played a key role in C&D’s experiment: participants had to choose between a limited range of options and the differences between the two versions of critical trials show that this had a significant effect. Hence, we could grant C&D that they found some embedded UBCs, chalk them up to contrastive construals, and go home. We could, but we’re not going to do that. Instead, we will argue that the most parsimonious interpretation of C&D’s data is that they didn’t find embedded UBCs of *any* kind, marked or unmarked.

One of the most striking aspects of C&D’s data is that the three pictures they used line up along a gradient: of the three, P1 is judged to fit the target sentence best, followed by P2, which in its turn is followed by P3. This gradient cannot be accounted for by supposing, as C&D do, that embedded UBCs were occasionally preferred, for though this would explain why P1 did better than P2 and P3, it would not explain why P2 received more votes than P3. Rather, we believe that the gradient is best explained by assuming that it is entirely due to a typicality effect.

The proposed diagnosis predicts that if we applied C&D’s paradigm with materials that are standardly used to demonstrate typicality effects, the pattern of results should mirror C&D’s. In order to test this prediction, we conducted the following experiment. We presented 26 native speakers of Dutch with items like the one in Figure 3, which had the same layout as in C&D’s study. In our case, the target sentence was (the Dutch equivalent of) *This is a bird*, and the two pictures showed one very typical and one rather untypical member of the class. The pattern of responses was the same as with the corresponding items in C&D’s experiment: while “both” was chosen 73% of the time, the more typical bird came in second with 24% of the votes, and the other two options were left to divide the remaining 3% between them.

Please indicate which picture is best described by the sentence below:

This is a bird.

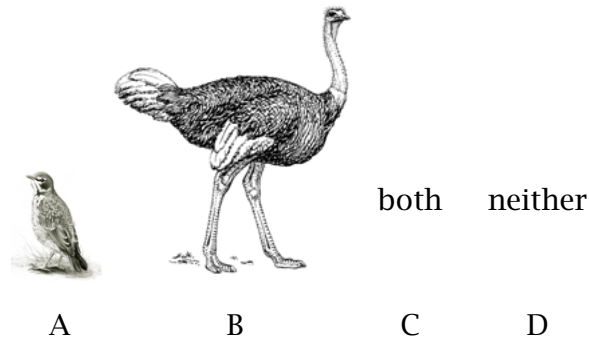


Figure 3: Critical item used in the typicality experiment.

Now the key observation is the following. It is clear that the reason why a substantial number of our participants voted for the robin is that it is a more typical bird than the ostrich. However, it patently does *not* follow that these participants interpreted the target sentence, *This is a bird*, as *meaning* that the individual in question is a robin. It is a well-known fact that typicality and class-membership are different things. An ostrich may be an atypical bird, but it is doubtlessly a bird. Perhaps the most striking illustration of the dissociation between typicality and class-membership is offered by mathematical concepts like “even number”. Experimental evidence shows that this concept, too, has typicality structure: two is the most typical even number (Armstrong, L. R. Gleitman & H. Gleitman 1983). But clearly, it would be nonsense to say that the preferred interpretation of (23) is that there are two tulips in the vase:

(23) The number of tulips in the vase is even.

By parity of reasoning, there is no need to suppose that embedded UBCs were implicated in C&D’s experiment: once we have invoked typicality to account for the gradient pattern in their data, there is nothing left for embedded UBCs to explain. Of course, it remains to be seen how the typicality structure associated with C&D’s critical sentence can be derived in a principled way. In the next section, we will outline how this can be done, but we

should like to emphasise that our argument doesn't hinge on that. For, on the one hand, we believe it is intuitively plausible that the gradient in C&D's data is a typicality effect, and on the other hand, our avian study proves that we can set up an experiment with the same design as C&D's that yields the same pattern of responses, which is best explained in terms of typicality, and typicality alone.

Summing up, there are two main problems with C&D's argument. One is that their experiment suffers from various design flaws, which undermine the conclusions they would like to draw. The other is that their data are better explained in terms of typicality than in terms of embedded UBCs. In the following we will argue that the same holds for Chemla & Spector's (2011) study.

5.2 Degrees of truth (Chemla & Spector 2011)

In Chemla & Spector's (2011) study, participants were presented with a rating task. Every item displayed a picture and a sentence, and participants had to indicate, on a continuous scale, how "true" or "appropriate" the sentence was, as a description of the picture. The critical sentence in C&S's experiment was:

(24) Every letter is connected to some of its circles.

This sentence was displayed with seven different pictures, three of which are shown in Figure 4. In each picture, a letter was connected with either all (A), none (N), or some but not all (M) of the circles surrounding it ("M" for "Mixed"). To illustrate, in Figure 4, P5 contains two M-cases and four A-cases, while P6 contains four M-cases and two A-cases.

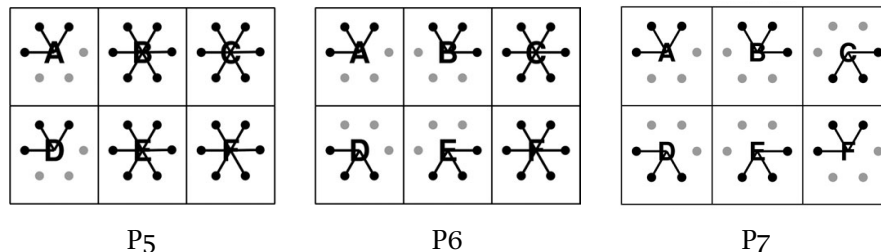


Figure 4: Pictures used in Chemla & Spector's (2011) study.

	P1	P2	P3	P4	P5	P6	P7
Number of N-cases	6	4	2	0	0	0	0
Number of A-cases	0	0	0	6	4	2	0
Number of M-cases	0	2	4	0	2	4	6
<i>Mean rating</i>	0	12	24	44	63	73	99

Table 2: All pictures used in critical items in Chemla & Spector’s (2011) experiment, with their mean ratings (%).

Table 2 gives the mean ratings for the experimental items in C & S’s study. In their interpretation of these data, C & S focus their attention on the contrast between the P5- and P6-conditions, on the one hand, and the P7-condition, on the other. Since P7 was the only picture that verified the target sentence with an embedded UBC, and its rating was significantly higher than those of the P5- and P6-items, which verified the sentence without an embedded UBC, C & S infer that their results support the existence of embedded UBCs. The obvious problem with this claim is that it accounts for part of the data only: like Clifton and Dube, C & S have a much richer data set than their experimental hypothesis can account for.

The astute reader will already see where we’re heading, but before we go there, let us first make a few preliminary points. For starters, C & S’s experimental instructions were verbose and less than fully coherent. Clocking in at well over 400 words (which would amount to about one page in this journal), they began by announcing that participants were to decide whether given sentences were true or false, but then this alethic dichotomy was qualified by the claim that, occasionally, target sentences “will not be clearly true or clearly false”, and henceforth the true/false terminology gave way to a variety of epithets, including (more or less) “natural”, “appropriate”, and even “true/appropriate”. (Chemla & Spector 2011: 392) In short, C & S’s instructions defy a straightforward construal, and there is no saying whether their participants achieved a consensual interpretation at all.

In Clifton and Dube’s study, it was hard to overlook the fact that contrasts between pictures were instrumental in shaping response patterns. Although at first sight, this doesn’t seem to hold for C & S’s experiment, two features of their design led us to suspect that contrast might have played a role in this case, too. First, participants saw every picture up to four times. Secondly,

C&S presented their critical items in one continuous series, not bothering to use fillers to disguise the purpose of the experiment. These two features could have facilitated and even invited comparisons between items. If that is what happened, how could contrastiveness have affected C&S's results? One possibility that suggests itself is that participants judged the key sentence less appropriate for a given picture only if they had previously encountered a better instance. To test this hypothesis, we compared participants' ratings in the P5- and P6-conditions preceding any P7-items with the same conditions occurring after participants had seen one or more P7-items. Sure enough, P5- and P6-items received higher ratings prior to P7, and the mean rating of the P6-items preceding P7 was statistically indistinguishable from that of the P7-items, both being close to 100%.

	P5	P6
Trials in C&S's experiment following P7	63	73
Trials in C&S's experiment preceding P7	78	96
Our replication	93	100

Table 3: Mean ratings (in %) for conditions P5 and P6 in C&S's experiment and our partial replication.

In order to corroborate this finding, we conducted a partial replication of C&S's experiment, in which we left out P7 entirely, and P5 and P6 were the only critical pictures. After three warm-up items, each participant saw the target sentence with either P5 or P6. In this experiment, there was no difference at all between the P5- and P6-conditions and C&S's P7-condition (Table 3). Therefore, what C&S present as the main effect in their data, and the keystone in their case for embedded UBCs, is an experimental artifact; more precisely, it is a contrast effect. First appearances notwithstanding, C&S's experiment is not very different from Clifton and Dube's.

As in the case of Clifton and Dube's, we claim that C&S's data, too, are best explained in terms of typicality, and that this provides us with an explanation that obviates the need for postulating embedded UBCs. Now, C&S admit that some of the regularities in their data may be due to typicality effects, but they argue that, even if they are, embedded UBCs are still needed:

[If] the main factor explaining these results is the one hypothesized by the "typicality interpretation", what must the underlying

ing metric be? More specifically, what kind of situations must be counted as “prototypical” instances of the sentence? As far as we can see, one should conclude that the best instances of the sentence among our various pictures are the ones used in the condition that receives the highest rating [i.e., P7] We would thus be led to conclude that the best instances of the sentence are those that make the [embedded UBC] true. But it is hard to see how this could be so if the local reading did not correspond to a salient reading of the sentence [...] So the “typicality interpretation”, as far as we can see, would support our conclusion that the [embedded UBC] exists. (Chemla & Spector 2011: 381-382)

To see why this is a non sequitur, recall from §5.1 that, even though two is the prototypical even number, (25) does not imply in any way that there are two tulips in the vase:

(25) The number of tulips in the vase is even.

Or in C&S’s words: there is no “salient reading” of (25) which says that there are two tulips in the vase. Similarly, if among C&S’s pictures, P7 is the best instance of (24), then it does not follow that this sentence has a salient reading according to which no letter is connected to all of its circles. So C&S’s argument is invalid.

Still, it remains to be seen how we can account for the typicality effects associated with multiply quantified sentences in a way that fits C&S’s data. The following analysis is developed in detail by van Tiel (to appear), so we will confine ourselves to the key ideas, starting with a concrete example. Imagine 11 situations (s_0, \dots, s_{10}), each of which contains 10 circles. In every s_n , n circles are black, and the remaining circles are white. How well does any s_n fit the following statements?

- (26) a. Every circle is black.
b. Some of the circles are black.

In order to answer this question, van Tiel asked 30 native speakers of English for their judgments, using a 7-point scale; Figure 5 gives the mean ratings he obtained for all situations (which were presented graphically) and each of the two sentences in (26).

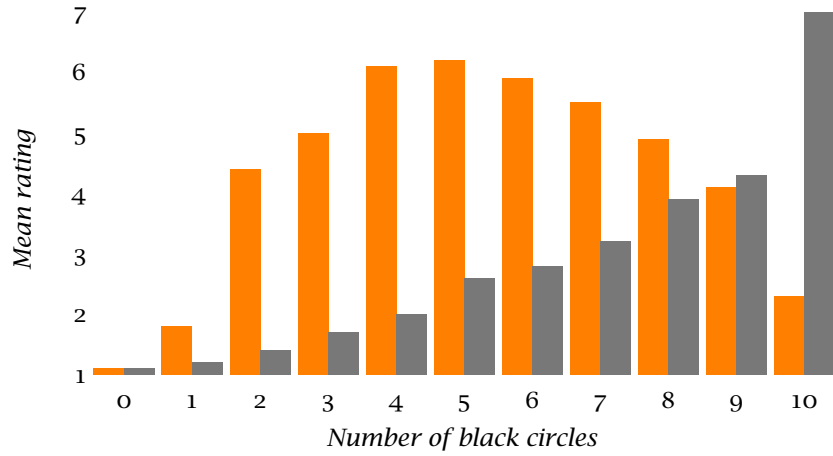


Figure 5: Mean ratings for (26a) (■) and (26b) (■) in situations consisting of ten circles, up to ten of which were black, while the rest were white.

Clearly, both sentences induce a distinctive prototype structure on the set $\{s_0, \dots, s_{10}\}$; that is to say, an ordering in terms of similarity to a “best case”. Whereas for (26a), s_{10} is the prototype, the best instances of (26b), according to van Tiel’s informants, are the situations in the middle region of the graph. In addition to similarity to a prototype, speakers’ judgments are influenced by the sentence’s truth value: situations in which a sentence is true received significantly higher ratings than situations in which it is false. This explains why s_{10} is the unique prototype for (26a): it is the only situation in which that sentence is true.

In van Tiel’s analysis, these observations are captured as follows. Let φ be a quantified sentence and s a situation consisting of a number of individuals $\{x_0, \dots, x_n\}$, each of which instantiate φ ’s predicate \mathcal{P} to a greater or lesser degree. $\tau_{\mathcal{P}}(x)$ and $\tau_{\varphi}(s)$ are numerical values that reflect how well x and s instantiate \mathcal{P} and φ , respectively. For universal sentences, $\tau_{\varphi}(s)$ could be analysed simply as the mean of the typicality values of the individuals in the contextual domain:

$$(27) \quad \tau_{\varphi}(s) := \frac{\sum \tau_{\mathcal{P}}(x_i)}{|s|}, \text{ where } s = \{x_0, \dots, x_n\} \text{ and } 0 \leq i \leq n$$

This is already a quite reasonable model of the pattern of ratings shown in Figure 5, but we can do better. Van Tiel’s data show that bad instances exert more influence on the mean rating of a situation than good instances:

whereas the situation with only one white circle received a significantly lower rating for (26a) than the prototype, there was almost no difference between the situations with one black circle and none. This observation can be modelled by weighing the typicality values of the individual instances in such a way that lower typicality values carry more weight than higher ones, which can be done simply by taking the harmonic instead of the arithmetic mean:

$$(28) \quad \tau_{\varphi}(s) := \frac{|s|}{\sum \tau_{\varphi}(x_i)^{-1}}, \text{ where } s = \{x_0, \dots, x_n\} \text{ and } 0 \leq i \leq n$$

Van Tiel shows that, for a very broad range of values for its free parameters, this analysis yields a strong correlation with the ratings he obtained for (26a) when paired with situations consisting of black and white circles.

For existential sentences, van Tiel proposes a coarse-level analysis that abstracts away from the typicality values of the individuals a situation consists of:

$$(29) \quad \tau_{\varphi}(s) := 1 - d_{\varphi}(s)^2 + v_{\varphi}(s)$$

Here, $d_{\varphi}(s)$ is the distance between s and the prototype situation, which according to van Tiel's informants was the situation consisting of an equal number of black and white circles. $v_{\varphi}(s)$ is a function of φ 's truth value in s , which can vary according to how strongly truth is hypothesised to influence typicality judgements. Hence, (29) is a straightforward implementation of the observation that $\tau_{\varphi}(s)$ is determined by its similarity to the prototype and φ 's truth value in s . This analysis, too, fits the observed pattern of ratings quite well for a very broad range of values for $v_{\varphi}(s)$.

Armed with the analyses in (28) and (29), we return to Chemla and Spector's experiment. As discussed at the beginning of this section, C&S presented sentence (30) paired with seven pictures (P1, ..., P7) containing six letters (ℓ_1, \dots, ℓ_6), each of which was connected to all, none, or some but not all of the circles surrounding it (Table 2):

(30) Every letter is connected to some of its circles. (=24)

According to the analysis in (28), the typicality value of any picture P is a function of the typicality values of its letters with respect to the predicate

(31) ... is connected to some of its circles.

For each picture P with letters ℓ_1, \dots, ℓ_6 , this yields:

$$(32) \quad \tau_{(30)}(P) = \frac{6}{\sum \tau_{(31)}(\ell_i)^{-1}}, 1 \leq i \leq 6$$

The typicality analysis of existential sentences in (29) entails the following:

- (33) Whenever ℓ is connected to none of its circles, ℓ' is connected to all of its circles, and ℓ'' is connected to some but not all of its circles, then $\tau_{(31)}(\ell) < \tau_{(31)}(\ell') < \tau_{(31)}(\ell'')$.

(32) and (33) constrain, but don't fix, the possible ratings associated with P_1, \dots, P_7 , and in order to determine how well these constraints capture C&S's data, van Tiel ran a Monte Carlo simulation, which involved generating 5,000 triplets of values satisfying the constraint in (33). For each triplet, (32) was used to compute the predicted typicality value for each of P_1, \dots, P_7 , and then the means of these values were compared with the mean ratings C&S found for these situations. The correlation was remarkably close to 1 ($r = .995, p < .001$). Hence, van Tiel's typicality model fits C&S's findings almost perfectly, and unlike C&S's own proposal, it accounts for the data practically without remainder.

To sum up: C&S's data exhibit the kind of gradient pattern that, even by their own admission, calls for an explanation in terms of typicality. Nevertheless, C&S insist that a small part of this pattern proves the existence of embedded UBCs. We have seen that C&S's argument cuts no ice: their findings can be fully accounted for by a typicality model that is entirely motivated on independent grounds, which obviates the need for embedded UBCs.

6 Non-monotone environments

Thus far, we have confined our attention to sentences in which a scalar expression occurs in an upward-entailing context; in this section, we turn to non-monotone environments:

- (34) There was only one key that fit some of the locks.

On some versions of the conventionalist account, (34) should "systematically and freely" give rise to the following reading:

- (35) There was only one key that fit some but not all of the locks.

Embedded scalars

- ≡ One key fit some but not all of the locks, and all the others fit either none or all of the locks.

Is this reading available in principle? Sure it is. We only have to apply contrastive stress to *some*:

(36) There was only one key that fit **SOME** of the locks.

Is this reading freely available? No: it evidently needs to be forced. That, at any rate, is our intuition, and it was supported by Geurts and Pouscoulous's data. In the study reviewed in §3, they presented sentence (37) with the pictures in Figure 6:

(37) There are exactly two circles that are connected with some of the squares.

Whereas the picture on the left makes the sentence true on the critical reading, the picture on the right makes it false. Geurts and Pouscoulous's first experiment robustly failed to corroborate conventionalist predictions: participants' truth-value judgments contradicted the predicted reading on every single trial. In their ambiguity study, Geurts and Pouscoulous found

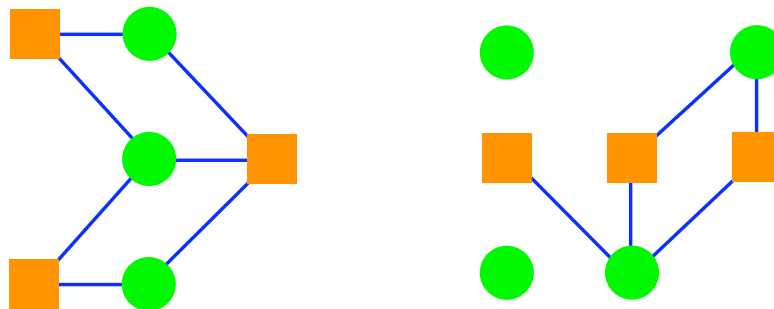


Figure 6: Pictures used by Geurts and Pouscoulous with sentence (37). Note that, with an embedded UBC for *some*, the picture on the left verifies this sentence, while the picture on the right falsifies it.

a modicum of support for conventionalists of the minimalist persuasion (cf. §3), but the evidence was quite weak: in this experiment, participants detected an ambiguity on 11% of the trials, which is better than nothing, but

uncomfortably close to nothing when compared to the 70% rate at which bona fide ambiguities were detected in control items.

As one might expect, Chemla and Spector claim that their participants did much better. Figure 7 shows two of the critical pictures in C&S's second experiment; the target sentence was the French equivalent of:

(38) There is exactly one letter connected with some of its circles.

C&S report that, in the critical condition, in which (38) was paired with the B-picture in Figure 7, the sentence received an average rating of 73%, which causes them to leap to the conclusion that there may be a *general* preference for deriving UBCs, be they embedded or not (Chemla & Spector 2011: 387).

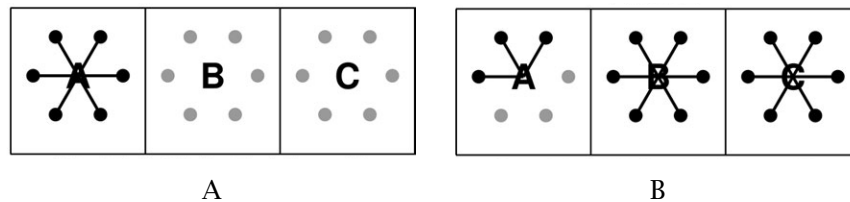


Figure 7: Pictures used by Chemla and Spector with sentence (38). Note that, with an embedded UBC for *some*, the picture on the left falsifies this sentence, while the picture on the right verifies it.

Apart from the fact that this conclusion would be premature even if C&S's evidence was sound, it is doubtful that the evidence they present *is* sound. To begin with, the same worries we voiced about C&S's first experiment apply in this case, too: the instructions were the same, hence susceptible to the same criticism (cf. §5.2), and as in the first experiment, the target sentence was shown over and over again to a small number of participants ($n = 16$), with no fillers between critical items. In short, it's almost as if C&S's experiment was designed to cause interactions between items. We have seen that this is what happened in their first experiment, and we will see that it happened in their second experiment, too.

In addition to potential contrast effects between items, it is evident that there was a salient contrast *within* the B-picture in Figure 7, too, and it is not hard to see how this might explain the high rating C&S obtained for this particular item. Given that, in this picture, there is a striking visual contrast between the leftmost letter and the other two, and given that the predicate of (38) is unequivocally verified by the left-most letter, informants

were invited to interpret that letter as the one the author of (38) had in mind, and construe the predicate accordingly; that is, interpret it in such a way that it was falsified by the remaining letters.¹⁰ Put otherwise, we maintain that the visual contrast in the B-picture provoked a contrastive interpretation of the predicate of the target sentence. If this is correct, C&S's data have no bearing on the debate between Gricean and conventionalist theories of embedded UBCs.

In a sense, C&S seem to agree with this assessment. Recall from §4 that C&S complained that Geurts and Pouscoulous's pictures were "difficult to decipher" (p. 365), and one of the virtues they claim for their own materials is that they are better in this respect. Now compare C&S's pictures in Figure 7 with Geurts and Pouscoulous's pictures in Figure 6. If C&S's claim is that the latter are more complex than the former, then it is debatable at best. However, if C&S's claim is that their pictures forced the reading they were after, while Geurts and Pouscoulous's didn't, then we fully agree. But in that case there is no issue.

Our diagnosis of C&S's main finding is that it is driven by a salient visual contrast in the B-picture in Figure 7. This assumption entails that if the contrast is reduced, the picture's rating should follow suit. In order to test that prediction, we conducted an experiment with materials that were nearly identical to C&S's (Table 4), but in which high-contrast and low-contrast critical items were presented to different groups of participants. The crucial difference between the high-contrast and low-contrast pictures was that in the former but not the latter, there was a stark visual contrast between one letter, which unequivocally verified the predicate of the target sentence in (39), and the remaining letters.

(39) Exactly one letter is connected to some of its circles.

Questionnaires for 54 participants were posted on Amazon's Mechanical Turk. All participants saw the same experimental items, with one exception only: whereas for half of the participants (Group A), sentence (39) was paired with the low-contrast picture (LOCON) in Table 4, for the other half (Group B), the same sentence was presented with the high-contrast picture (HICON), i.e. C&S's B-picture. The remaining three items were the same for both groups;

¹⁰ Here we assume, with C&S, that hearers adopt the Principle of Charity by default. That is to say, we accept the *original* Principle of Charity (16), which we consider to be plausible, in contrast to C&S's Preference for Truth (18), which is much stronger, and not plausible at all (see §4).

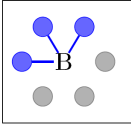
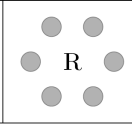
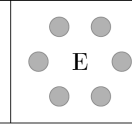
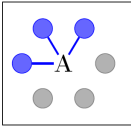
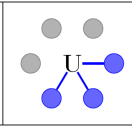
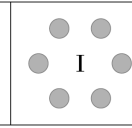
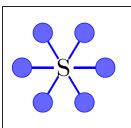
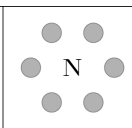
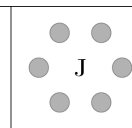
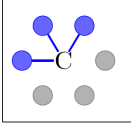
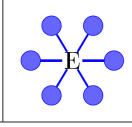
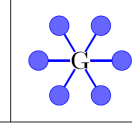
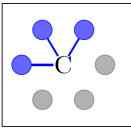
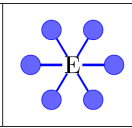
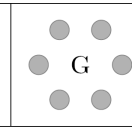
TRUE				plain true
FALSE				plain false
LITERAL				false with an embedded UBC
HiCON				true with an embedded UBC high contrast
LOCON				true with an embedded UBC low contrast

Table 4: Experimental items used in the in the high/low contrast study.

they featured pictures in which the target sentence was either unequivocally false (FALSE), unequivocally true (TRUE), or only literally true, i.e. false with an embedded UBC for *some* (LITERAL). The order of presentation was randomised across 9 lists. As in C&S's experiment, we didn't include any fillers in our experiment, so there were four items in both conditions. In comparison to C&S's instructions, ours were short and sweet: participants were simply asked to indicate, on a 7-point scale, how well each of the pictures was described by sentence (39).

Three participants were excluded from the analysis because they weren't native speakers of English. The average ratings produced by the remaining participants are given in Table 5. As expected, the results for the high-contrast condition (Group B) mirror C&S's. Participants considered (39) the best description of the TRUE picture, followed by the HiCON, LITERAL, and FALSE pictures (in that order). Most importantly, the difference between the HiCON and FALSE items was statistically significant ($t(25) = 4.035$, $p < .001$). In the low-contrast condition (Group A), on the other hand, the ratings for the LOCON and FALSE items were not significantly different ($t(24) = 1.252$, $p =$

	Group A	Group B
TRUE	84	80
FALSE	30	32
LITERAL	74	49
HiCON	-	64
LoCON	37	-

Table 5: Average ratings (in %) in the high/low contrast experiment.

.22). Hence our prediction was confirmed: it appears that the high rating for the HiCON item (= C & S’s B-item) was due to a visual contrast within that picture.¹¹

The upshot of the foregoing discussion is that the experimental record offers no support to the claim that scalar items freely license upper-bounded construals in non-monotone environments. In the last section, we argued that the same holds for upward-entailing environments. Hence, we conclude that, as things currently stand, there is no experimental evidence in favour of the view that scalar expressions freely give rise to embedded UBCs.

Conclusion

It is widely agreed that there are two mechanisms that variously underwrite UBCs of scalar expressions: conversational implicature and truth-conditional narrowing. The main bone of contention in the debate on so-called “embedded implicatures” concerns the division of labour between these two mechanisms. More precisely, it is about the availability of truth-conditional narrowing. On the conventionalist view, narrowing is freely available, and therefore we

¹¹ Interestingly, C & S also tested both HiCON and LoCON pictures in their experiment, but failed to find a difference between them: both received a comparatively high rating of 73%. A likely explanation for the discrepancy between C & S’s results and ours is that, once participants had encountered a HiCON item, the UBC triggered by this item had knock-on effects on subsequent items. This would also explain why, in our experiment, the difference between the LITERAL and TRUE items was significant for Group B ($t(25) = 4.1, p < .001$), but only marginally so for Group A ($t(24) = 1.959, p = .062$). In C & S’s study, such effects may have been reinforced by the extreme repetitiveness of the experiment, in which every item was presented four times, which is to say that each participant saw the critical sentence no less than 52 times.

should expect to regularly observe its effect in embedded positions, too. On the Gricean view, narrowing requires special circumstances, like contrastive stress for example, and therefore embedded UBCs should be the exception rather than the rule.

This issue cannot be decided on the basis of introspective evidence, since introspection can only tell us *if* a given expression can be interpreted in such-or-such a way, and there is a consensus that, e.g., (40a) *can* be interpreted as (40b):

- (40) a. All the squares are connected with some of the circles.
 b. All the squares are connected with some but not all of the circles.

The key question is whether such readings are readily available, and we can only hope to answer that question by collecting experimental data. Obviously, experiments should be designed in such a way that they don't bias participants towards the contested reading, for again, it is agreed that embedded UBCs are possible: what needs to be established is whether or not they "occur systematically and freely in arbitrarily embedded positions", as Chierchia, Fox & Spector (2012) put it.

A recurrent flaw in the experimental studies that have been claimed to support the conventionalist position is that they did bias participants towards the desired response. This was done (unwittingly, of course) by means of contrasts between or within items, the effect of which was reinforced, in Chemla and Spector's study, by the way materials were presented. As a result, the experimental findings reported by Clifton and Dube and Chemla and Spector fail to prove their point. Fortunately, however, their data are still of great interest, since they show that embedded scalars may give rise to typicality effects. This is an important result, and the fact that it was serendipitous doesn't diminish its importance.

To return to our main topic, we conclude that as far as the experimental evidence is concerned, the state of the art is where Geurts and Pouscoulous (2009) left it, which is to say that the data speak unequivocally against conventionalism and in favour of a Gricean view on scalar expressions.

References

- Armstrong, Sharon Lee, Lila R. Gleitman & Henry Gleitman. 1983. What some concepts might not be. *Cognition* 13. 263–308. [http://dx.doi.org/10.1016/0010-0277\(83\)90012-4](http://dx.doi.org/10.1016/0010-0277(83)90012-4).

- Bach, Kent. 1994. Conversational implicature. *Mind and Language* 9. 124-162. <http://dx.doi.org/10.1111/j.1468-0017.1994.tb00220.x>.
- Bartsch, Renate. 1987. Context-dependent interpretations of lexical items. In Jeroen Groenendijk, Dick de Jongh & Martin Stokhof (eds.), *Foundations of pragmatics and lexical semantics*, 1-26. Dordrecht: Foris.
- Bosch, Peter. 1984. *Lexical learning, context dependence, and metaphor*. Indiana University Linguistics Club.
- Chemla, Emmanuel & Benjamin Spector. 2011. Experimental evidence for embedded scalar implicatures. *Journal of Semantics* 28. 359-400. <http://dx.doi.org/10.1093/jos/ffq023>.
- Chierchia, Gennaro. 2004. Scalar implicatures, polarity phenomena and the syntax/pragmatics interface. In Adriana Belletti (ed.), *Structures and beyond*, 39-103. Oxford University Press.
- Chierchia, Gennaro. 2006. Broaden your views: Implicatures of domain widening and the “logicality” of language. *Linguistic Inquiry* 37. 535-590. <http://dx.doi.org/10.1162/ling.2006.37.4.535>.
- Chierchia, Gennaro, Danny Fox & Benjamin Spector. 2012. Scalar implicature as a grammatical phenomenon. In Claudia Maienborn, Klaus von Stechow & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning, volume 3*, 2297-2332. Mouton de Gruyter.
- Clifton, Charles & Chad Dube. 2010. Embedded implicatures observed: A comment on Geurts and Pouscoulous (2009). *Semantics and Pragmatics* 3. 1-13. <http://dx.doi.org/10.3765/sp.3.7>.
- Cohen, L. Jonathan. 1971. The logical particles of natural language. In Yehoshua Bar-Hillel (ed.), *Pragmatics of natural language*, 50-68. Dordrecht: Reidel.
- Davidson, Donald. 1973. Radical interpretation. *Dialectica* 27. 314-328. <http://dx.doi.org/10.1111/j.1746-8361.1973.tb00623.x>.
- Fox, Danny. 2007. Free choice and the theory of scalar implicatures. In Uli Sauerland & Penka Stateva (eds.), *Presupposition and implicature in compositional semantics*, 71-120. Houndmills, Basingstoke: Palgrave Macmillan.
- Geurts, Bart. 2000. Buoyancy and strength. *Journal of Semantics* 17. 315-333. <http://dx.doi.org/10.1093/jos/17.4.315>.
- Geurts, Bart. 2009. Scalar implicature and local pragmatics. *Mind and Language* 24. 51-79. <http://dx.doi.org/10.1111/j.1468-0017.2008.01353.x>.
- Geurts, Bart. 2010. *Quantity implicatures*. Cambridge University Press.
- Geurts, Bart & Nausicaa Pouscoulous. 2009. Embedded implicatures?!? *Semantics and Pragmatics* 2. 1-34. <http://dx.doi.org/10.3765/sp.2.4>.

- Grice, Herbert Paul. 1975. Logic and conversation. In Peter Cole & J. Morgan (eds.), *Syntax and semantics 3: Speech acts*, 41–58. New York, NY: Academic Press.
- Horn, Laurence R. 1989. *A natural history of negation*. Chicago University Press.
- Horn, Laurence R. 2006. The border wars: A neo-Gricean perspective. In Klaus von Heusinger & Ken Turner (eds.), *Where semantics meets pragmatics*, 21–48. Amsterdam: Elsevier.
- Horn, Laurence R. 2009. WJ-40: Implicature, truth, and meaning. *International Review of Pragmatics* 1. 3–34. <http://dx.doi.org/doi:10.1163/187731009X455820>.
- Ippolito, Michela. 2010. Embedded implicatures? Remarks on the debate between globalist and localist theories. *Semantics and Pragmatics* 3. 1–15. <http://dx.doi.org/10.3765/sp.3.5>.
- Landman, Fred. 1998. Plurals and maximalization. In Susan Rothstein (ed.), *Events and grammar*, 237–271. Dordrecht: Kluwer.
- Landman, Fred. 2000. *Events and plurality: The Jerusalem lectures*. Dordrecht: Kluwer.
- Levinson, Stephen C. 2000. *Presumptive meanings*. Cambridge, MA: MIT Press.
- Magri, Giorgio. 2009. A theory of individual-level predicates based on blind mandatory scalar implicatures. *Natural Language Semantics* 17. 245–297. <http://dx.doi.org/10.1007/s11050-009-9042-x>.
- Meyer, Marie-Christine & Uli Sauerland. 2009. A pragmatic constraint on ambiguity detection. *Natural Language and Linguistic Theory* 27. 139–150. <http://dx.doi.org/10.1007/s11049-008-9060-2>.
- Noveck, Ira & Dan Sperber. 2007. The why and how of experimental pragmatics: The case of “scalar inferences”. In Noel Burton-Roberts (ed.), *Pragmatics*, 184–212. Basingstoke: Palgrave Macmillan.
- Nunberg, Geoffrey D. 1978. *The pragmatics of reference*. Indiana University Linguistics Club.
- Nunberg, Geoffrey D. 1979. The non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy* 3. 143–184. <http://dx.doi.org/10.1007/BF00126509>.
- Quine, Willard Van Orman. 1960. *Word and object*. Cambridge, MA: MIT Press.
- Recanati, François. 1986. On defining communicative intentions. *Mind and Language* 1. 213–241. <http://dx.doi.org/10.1111/j.1468-0017.1986.tb00102.x>.

- Recanati, François. 2003. Embedded implicatures. *Philosophical Perspectives* 17. 299-332. <http://dx.doi.org/10.1111/j.1520-8583.2003.00012.x>.
- Recanati, François. 2004. *Literal meaning*. Cambridge University Press.
- van Rooij, Robert & Katrin Schulz. 2004. Exhaustive interpretation of complex sentences. *Journal of Language, Logic and Computation* 13. 491-519. <http://dx.doi.org/10.1007/s10849-004-2118-6>.
- Russell, Benjamin. 2006. Against grammatical computation of scalar implicatures. *Journal of Semantics* 23. 361-382. <http://dx.doi.org/10.1093/jos/ffloo8>.
- Sag, Ivan. 1981. Formal semantics and extralinguistic content. In Peter Cole (ed.), *Radical pragmatics*, 273-294. New York, NY: Academic Press.
- Sauerland, Uli. 2004. Scalar implicatures in complex sentences. *Linguistics and Philosophy* 27. 367-391. <http://dx.doi.org/10.1023/B:LING.0000023378.71748.db>.
- Sauerland, Uli. 2010. Embedded implicatures and experimental constraints: A reply to Geurts & Pouscoulous and Chemla. *Semantics and Pragmatics* 3. 1-13. <http://dx.doi.org/10.3765/sp.3.2>.
- Spector, Benjamin. 2006. *Aspects de la pragmatique des opérateurs logiques*. Doctoral dissertation, University of Paris VII.
- Storto, Gianluca & Michael K. Tanenhaus. 2005. Are scalar implicatures computed online? *Sinn und Bedeutung* 9. 431-445.
- van Tiel, Bob. to appear. Embedded scalars and typicality. *Journal of Semantics*. <http://dx.doi.org/10.1093/jos/fft002>.
- Wilson, N. L. 1959. Substances without substrata. *The Review of Metaphysics* 12. 521-539. <http://www.jstor.org/stable/20123725>.

Bart Geurts
Department of philosophy
University of Nijmegen
Postbus 9103
6500 HD Nijmegen
The Netherlands
brtgrts@gmail.com

Bob van Tiel
Department of philosophy
University of Nijmegen
Postbus 9103
6500 HD Nijmegen
The Netherlands
bobvantiel@gmail.com