Distinguishing typicality and ambiguities, the case of scalar implicatures

Emmanuel Chemla and Benjamin Spector[1]

Abstract

We refine previous methods to evaluate the existence of local scalar implicatures. We show that these refined methods further confirm the existence of local scalar implicatures (contra previous claims). We also demonstrate that these same methods can be applied to settle other controversies, not restricted to scalar implicatures. Specifically, these methods provide the way to quantify the relative salience of different interpretations for ambiguous sentences, generally. Based on this method, we defend a conceptually motivated theory of 'typicality projection'.

1 Introduction: truth-conditions, scalar implicatures and typicality

Formal semantics often targets categorial data: is such and such sentence true/felicitous or not in a given situation? The access we have to such judgments, whether we use introspective or more quantitative methods, may however be easily blurred by other aspects of language and language use, as well as the nature of the judgment task itself. For instance, in a sentence-picture matching task (where participants are asked to assess whether a sentence correctly describes a picture), the participants's responses may well reflect much more than just their truth-conditional intuitions. Vagueness and ambiguities, as well as specific aspects of the judgment task itself, may blur the binary true/false picture. It follows that in order to draw firm conclusions from a truth-value judgment task regarding a given theoretical question, one needs to have in mind, on top of a clear theoretical hypothesis, a testable model of the task itself, i.e. precise hypotheses as to the way non truth-conditional aspects affect subjects' performance.

This paper is a case study in which we investigate simultaneously a specific theoretical hypothesis and the properties of a specific truth-value judgment task. We will be concerned with two phenomena whose interaction may blur truth-conditional intuitions: *typicality* (understood in a broad sense) and *scalar implicatures*.

Typicality refers to the attraction to some specific exemplars of a kind, which may count as better representatives of their kind: a sparrow is a more typical exemplar of a bird than a penguin. By extension, we will use the word 'typicality' to refer to the attraction to some situations, as more or less good representatives of (the truth of) a proposition. In that sense, typicality could surely blur categorial truth-value judgments. One may think that typicality plays a particularly significant role when the judgment task involves a graded scale (rather than binary judgments), but in principle it might well play a role in binary judgments as well.

A typical instance of a scalar implicature is the inference from a sentence such as (1-a) to the conclusion that (1-b) holds.

(1)   a.   John ate some of the cookies.
           b.   John did not eat all of the cookies.

The classical view is that this inference, which appears to be optional, is not a logical inference, i.e. does not directly follow from the truth-conditions of (1-a) but rather results from a reasoning about speakers' communicative goals. However, scalar implicatures have given rise to a debate as to whether they are really pragmatic in nature or rather reflect a genuine semantic ambiguity (see Chierchia et al., 2012 and the references cited therein for a defense of the latter view). Whether they are related to pragmatic, flexible processes or are a specific type of semantic ambiguity, scalar implicatures may interact with and may be hard to distinguish from other phenomena of a hybrid kind, e.g., typicality. This creates both conceptual and methodological difficulties if one wants to settle empirical controversies about scalar implicatures.

The starting point of the current paper is a controversy in the realm of scalar implicatures. The short version of the history is the following: some have argued for the existence of local scalar implicatures (see Chierchia et al., 2012 and the references cited therein). Geurts and Pouscoulous (2009) have produced data in which there was no sign of these inferences; Clifton and Dube (2010) and Chemla and Spector (2011) have provided counterevidence arguing for the existence of local scalar implicatures; van Tiel (2013) has argued that these studies were confounded by typicality issues.

In this paper, we first discuss the current situation with respect to the localism vs. globalism debate (section 2). In essence, we show that van Tiel (2013) offers a powerful method to re-assess Chemla and Spector's results, but that once it is applied properly and without eliminating arbitrarily half of the data, the result is a confirmation of the initial interpretation. We will then take advantage of van Tiel's (2013) insights in two ways. On the methodological side, we will show that we now have the means to quantify the relative salience of different readings for a sentence (section 3). On the more theoretical side, we further discuss the 'typicality structure' of the quantifiers *some* and *every*, as an interesting phenomenon *per se*, and offers a model which can lead to a theory of typicality projection (section 4).

2 Scalar implicatures: assessment of the quantitative arguments with van Tiel's method

2.1 Theories of scalar implicatures: globalism vs. localism

According to the Gricean approach to scalar implicatures (SIs for short), SIs are pragmatic inferences that result from a reasoning about the speaker's communicative intentions. In recent years, an alternative view of SIs (let us call it the 'grammatical view' of SIs) has been put forward, according to which they result from the optional presence of a covert so-called exhaustivity operator in the logical form of the relevant sentences and are thus reducible to semantic entailment. (see Chierchia, 2006; Fox, 2007, Chierchia et al., 2012, a.o, building on earlier grammatical approaches by, e.g., Landman, 1998, Chierchia, 2004).

While these two radically different approaches do not make distinct predictions in simple cases, they do for more complex ones. In particular, if the 'grammatical approach' is correct, then the exhaustivity operator should be able to occur in an embedded position (just like *only*), so that the strengthening, say, of *some* into *some but not all* could occur 'locally', under the scope of linguistic operators. This approach is often called 'localist', as opposed to more pragmatic, so-called 'globalist' approaches.

Consider for concreteness the following example:

(2)   Every student solved some of the problems.

The standard neo-Gricean mechanism predicts that (2) should be interpreted as implying the negation of its scalar alternative, i.e. the negation of 'Every student solved all of the problems'. Hence, (2) should give rise to the following reading (henceforth, we will refer to this reading as the 'global reading'):

(3)   Every student solved some of the problems and at least one student didn't solve them all.

If, however, the strengthening of *some* into *some but not all* can occur at an embedded level, as predicted by localist approaches, one expects that another possible reading for (2) is the one expressed by (4) below (which we will henceforth call the 'local reading'):

(4)   Every student solved some but not all the problems.

It thus seems that determining the possible readings of sentences like (2) should provide decisive evidence in the debate between localism and globalism. This is unfortunately not so. For several formalized *globalist* theories of SIs (e.g., Spector, 2003, 2006; van Rooij and Schulz, 2004; Chemla, 2008, 2009b) also predict that (4) is a possible reading of (2).[2] Before we continue, we would like to point out that quite generally we will be looking at a small subset of the arguments that have been put forward in this vivid debate. It seems to us that quantitative arguments coming from experimental methods have gained prominence. However, to draw robust conclusions from these data, one must have a clear view of the theoretical landscape. Note, in particular, that embedding under *every* is not the best possible case, as was discussed in, e.g., Chemla and Spector (2011). Non-monotonic operators provide a much better test case.[3] We will stick to *every* here though, because this is where van Tiel's analysis is relevant. Contrary to van Tiel however, we will not also restrict ourselves to data concerning cases where *some* is embedded under *every*. The interpretation we will propose for our data generalizes to the cases where *or* is embedded under *every*, which van Tiel ignored.

2.2   Chemla and Spector's results

Geurts and Pouscoulous (2009) tested sentences such as (2), but failed to find evidence for local readings such as (4). Chemla and Spector (2011) used a similar sentence-picture matching task, and did detect local readings. One important difference between the two sets of experiments is that instead of asking for absolute judgments of truth or falsity, Chemla and Spector asked for graded judgments: subjects were asked to position a cursor on a continuous line going from 'No' (i.e. 'false') on the left, to 'Yes' (i.e. 'true') on the right (see Fig. 1).[4] Offering informants more options than just true or false may help capture more fine-grained results, which could reveal differences that remained hidden when subjects are given only two or three options. This manipulation, however, may also allow other, irrelevant phenomena to intrude into the results.

---

[2]These theories do not derive this reading by localist means, of course. They argue instead that the proposition: *Some students solved all the problems* should be added to the list of negated scalar alternatives of (2).

[3]Strikingly, it is hard to extend van Tiel's view to the results we obtained for cases where a scalar item is embedded under a non-monotonic operator such as *exactly one*, and no discussion of these cases is offered.

[4]See Chemla (2009a,c) for the use of a similar methodology to collect judgments in pragmatics, and the references cited therein.

Figure 1: Response choice in Chemla and Spector's (2011) experiments.

We will discuss this in more detail below, as this is the criticism put forward most prominently by van Tiel (2013).

Chemla and Spector tested sentences of the form:

(5) a. Every letter is connected with some of its circles. (EVERY-SOME-sentences)
    b. Every letter is connected with its blue or its red circle. (EVERY-OR-sentences)

They asked participants to evaluate such sentences when paired with different types of pictures containing six cells in which one letter was surrounded by circles. For the EVERY-SOME-sentence, let us call a cell a *verifier* if the letter in the cell is connected with some or all of its circles (respectively with its blue circle or its red circle or both in the case of EVERY-OR-sentences). Let us call a cell a *strong verifier* if the letter in the cell is connected with some but not all of its circles (respectively with its blue circle or its red circle but not both). Now, on its literal meaning, the sentence is true if and only if each cell is a verifier. On its 'globally strengthened' meaning,[5] the sentence is true if and only if each cell is a verifier and at least one cell is a strong verifier. On the 'local' reading, the sentence is true if and only if each cell is a strong verifier. The results are repeated in Fig. 2.
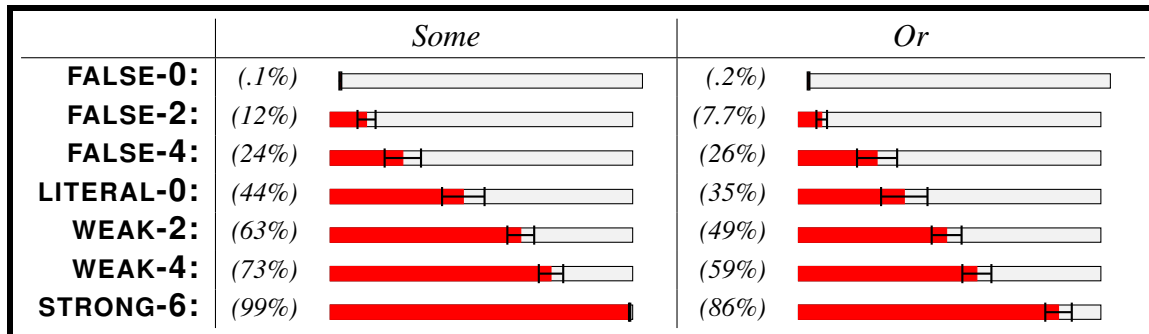

Figure 2: Results from Chemla and Spector (2011), including sub-conditions defined by the number of strong verifiers in the picture.

The code in the left column encodes the two relevant properties of the picture the sentence is paired with in a given condition: the word encodes the set of readings which were true and the number in this code indicates the number of strong verifiers in the picture.[6] We interpreted the fact that the **STRONG**-pictures (all readings true) were rated higher than the **WEAK**-pictures

---

[5]Regarding the difference between the 'global' reading and the 'local' reading, we refer the reader to our original paper.

[6]Specifically, in the **FALSE** cases: all readings were false (not every cell is a verifier); **LITERAL**: only the literal interpretation (everyone of the six cells is a verifier but no cell is a strong verifier) is true; **WEAK**: both the global and the literal readings are true but not the local reading (every cell is a verifier, some but not all of the cells are strong verifiers); **STRONG** all readings literal, global and local are true (all the six cells are strong verifiers).

(all readings *except the local reading* true) as evidence for the existence of the local reading. Overall, the results lend themselves to the following interpretation: the three hypothesized readings (literal reading, global reading and local readings) exist, and the more readings a picture makes true, the higher the sentence-picture pair is rated.

However, as we pointed out, the scores seem to vary not only depending on which set of readings is true, but also following the 'number of strong verifier' factor: even within the **FALSE** pictures, participants provided higher rates for pictures containing more strong verifiers (compare **FALSE-0**, **FALSE-2** and **FALSE-4**, but also **WEAK-2** and **WEAK-4**). This suggested to us the following refined interpretation of our results: when two pictures are evaluated relative to a certain reading, then even if both pictures make the reading false, the picture that is, in some informal sense, 'closer' to a case making the reading true will receive a higher score. The reason why **FALSE-4** receives a higher score than **FALSE-2** would then be that the **FALSE-4**-pictures are less different from, say a **WEAK** picture, than the **FALSE-2**-pictures are. Hence, the results would then be accounted by two interrelated factors:

(6)   Two components in Chemla and Spector's (2011) results:
   a.   Ambiguity component: which readings/how many readings are true?
   b.   Typicality component: how close is the picture to a true case?

In Chemla and Spector (2011), all the efforts were concentrated on the first component, (6-a) and no quantitative evaluation of the role of (6-b) was proposed. van Tiel (2013) on the contrary proposed a method to evaluate the typicality component (without assuming, as we did, that typicality reflects 'closeness to a true case') and concluded that the effect of (6-a) was minimal and that this factor was irrelevant to explain the difference between **WEAK-4** and **STRONG-6**. We will review van Tiel's argument below (see section 2.3). We will argue that what he proposes is a method to evaluate the contribution of the typicality factor, rather than a method to evaluate the absence of contribution of (6-a). We will then apply his method to demonstrate that there is an effect of (6-a), that is an effect of the local reading, beyond tipycality (see section 2.4).

2.3   van Tiel's challenge and method

van Tiel focussed on the EVERY-SOME case. He proposes an alternative interpretation of Chemla and Spector's data, one purely in terms of *typicality*. The view, more precisely, is that the predicate 'x is connected with some of its circles', even on its *literal* reading, has a certain *typicality structure*, whereby situations where x is connected with all of its circles count as less typical than situations where x is connected with, say, half of its circles. Then, depending on how typicality 'projects' under universal quantification, we may expect the **STRONG**-pictures to be the most 'typical instances' of the universally quantified sentence, independently of issues of implicature, because when each letter *typically* instantiates the predicate 'be connected to some of its circles', one may expect that the universal sentence in (5) turns out to be itself typically instantiated.

It might be argued that this notion of typicality is in fact derived from scalar implicatures (see discussion in section 4.1). But van Tiel reports experimental data which he takes to show that the 'typicality structure' of predicates containing 'some' does not just follow from the fact that it implicates 'not-all'. For instance, situations where five dots out of ten are black are

judged more 'typical' instances of the sentence 'some of the dots are black' than situations where two dots out of ten are black, and there is no *straightforward* explanation for this fact in terms of scalar implicature (a point we will return to in section 4). Based on experimental results regarding the typicality structure of 'some' and how typicality projects under 'every', van Tiel offered a model of our data in terms of typicality that does not involve scalar implicatures. His model is based on the two following assumptions:

(7) **Assumption 1- Typicality projection**. Consider a universally quantified sentence 'Every A is a B', in which B is a predicate subject to typicality. Let us then ask whether a certain situation s is a typical instantiation of the sentence. For each individual α_i, let us note $\rho_B(\alpha_i)$ the typicality value of s relative to the atomic sentence 'α_i is a B'. Then the typicality value of s relative to 'Every A is a B' is the *harmonic mean* of all the values $\rho_B(\alpha_i)$, where α_i ranges over all the A-individuals. In other terms, $\rho_{\text{every}}(s)$, the typicality value of s with respect to the universally quantified sentence, is given by:

$$\rho_{\text{every}}(s) = \frac{n}{\sum \frac{1}{\rho_B(\alpha_i)}}$$

(8) **Assumption 2 - Typicality structure of *some As are Bs***. Let us now call $\rho_{\text{some}}(\text{mixed})$ the typicality value of a situation in which some but not all As are Bs, relative to the sentence *some As are Bs*. Likewise, let $\rho_{\text{some}}(\text{all})$ be the typicality value of a situation in which all As are Bs, and $\rho_{\text{some}}(\text{none})$ that of a situation in which no A is B. Based on his results regarding the typicality structure of *some*, van Tiel is justified in assuming the following:

$$\rho_{\text{some}}(\text{mixed}) > \rho_{\text{some}}(\text{all}) > \rho_{\text{some}}(\text{none})$$

On this basis, van Tiel performed the following simulation:

(9) van Tiel's simulation: evaluating the role of typicality
   a. Pick randomly a triplet of ordered values for $\rho_{\text{some}}(\text{mixed})$, $\rho_{\text{some}}(\text{all})$, $\rho_{\text{some}}(\text{none})$, complying with Assumption 2.
   b. Compute the typicality value for a given picture, based on these 3 values and the formula given in Assumption 1, namely:

$$\rho_{\text{every}}(\text{picture}) = \frac{6}{\frac{n_{\text{mixed}}}{\rho(\text{mixed})} + \frac{n_{\text{all}}}{\rho(\text{all})} + \frac{n_{\text{none}}}{\rho(\text{none})}}$$

   with: – n_{mixed} the number of strong verifiers
         – n_{all} the number of cells in which the letter is connected to all its circles (i.e. non-strong verifiers)
         – n_{none} the number of cells where the letter is not connected to any circle.
   c. Do this 5,000 times and average over the results for each of the seven pictures.

The result is then estimated as the product-moment between these average scores, the 'predicted values', and the observed values (from Chemla and Spector, 2011). van Tiel obtains a correlation score of $r = .99, p < .001$, which shows that typicality as construed by van Tiel can

explain most of the observed pattern. van Tiel thus argues that Chemla and Spector's results can be entirely explained in terms of typicality, and therefore do not provide any evidence for the existence of the local reading.

However, there are two problems with this argument. First, granting, for the sake of the argument, that typicality as construed by van Tiel is totally independent of scalar implicatures, van Tiel's simulation simply does not address in any way the question whether the local reading exists. What it shows is that typicality can explain 98% of the data ($r^2 > .98$). This does not tell us whether taking into account on top of typicality the various (hypothetical) readings *improves* the fit of the simulation, i.e. adds explanatory power. But this, it seems to us, is in fact the critical question in this controversy. Second, in the absence of any explicit theory of typicality, we cannot exclude that van Tiel's measure of typicality may already reflect in part the role played by scalar implicatures. In our paper, we acknowledged the role played by typicality, but added that it seemed to us that "our results can be interpreted as reflecting typicality judgments *only if* typicality is construed as relative to *several* distinct readings." While this may have been too strong a claim, we will argue in section 4 that our initial conjecture is in fact extremely plausible, and preferable given certain considerations of parsimony and explanatory value.

2.4 A modified version of van Tiel's simulation: taking into account readings on top of typicality improves the fit

We had proposed that for a given sentence S, all things being equal, if the readings that are true in condition C1 properly include the readings that are true in condition C2, the sentence will receive a higher score in C1 than in C2. The picture however is blurred by typicality issues. Assuming that typicality as construed by van Tiel is totally independent of scalar implicatures (a very strong assumption that we will question in section 4), what we need to assess, then, is whether taking into account the truth-value of various readings *on top of typicality* considerations improves the overall fit of the simulation. To evaluate this, we inserted a step in the simulation, between the second and the third step. We first chose an arbitrary value V, namely $V = 10$,[7] for rewarding readings (recall that possible scores range from 0 to 100), and given a certain set of possible readings R, we proceeded as follows:

(10)   Simulation 1: evaluating the explanatory value of different sets of readings, beyond typicality

    a.   Pick randomly a triplet of ordered values for ρ(mixed), ρ(all), ρ(none), complying with Assumption 2. (same as above)

    b.   Compute the typicality value for a given picture, based on these 3 values and the formula given in Assumption 1. (same as above)

    c.   <u>Additional step</u>: For each picture, add to the typicality value a reward: V×(number of readings in R that the picture makes true). So, if no reading in R is true relatively to the picture, add nothing. If exactly one reading in R is true, add the reward V once, etc.[8]

---

[7]We have tried several values here and it does not change anything; in fact, we also tried changing this value randomly along with the rest of the sampling procedure and again the result was similar so we present the simple version here.

[8]We applied this step for each set of readings on the same triplets, which allows us to run *paired* tests to compare the contribution of different sets of readings.

d. Do this 100 times and average over the results for each of the seven pictures. (same as above, except for the number of iterations)

Different theories predict different sets of possible readings, and we can now compare the different possible sets of readings, by comparing the fit that each set leads to, and ask which set of readings provides the best fit (rather than which set provides a good fit). Let us note immediately that it is not the case that considering more readings (or readings at all) provides an intrinsic advantage to the models. The situation is not one in which we compare two linear models with one relying on a larger number of predictors than the other. Instead, each model makes a specific prediction, which may or may not be borne out.[9]

To compare the sets of readings on solid grounds, we used a bootstrapping technique: for each set of readings, we ran the procedure above 100 times to gather 100 scores obtained for each set of readings (a score is the product-moment correlation between the predicted values of the simulation and the observed values). We were thus in a position to compare the distribution of these scores.[10] Assuming that the literal and the global readings uncontroversially exist, we want to know which of the sets of readings $\mathcal{L}$(ocalism)={literal, global, local} and $\mathcal{G}$(localism)={literal, global} better predict the results. We also ran the same simulation for the EVERY-OR case, using the following typicality ordering for 'A or B': ρ_{or}(situations where A or B but not both are true) $>$ ρ_{or}(situations where both A and B are true) $>$ ρ_{or}(situations where neither A nor B is true). We find that $\mathcal{L}$ offers a significantly better predictor than $\mathcal{G}$, both when considering the *some* sentences (.998 vs .993, $t(99) = 20$, $p < 10^{-15}$) and the *or* sentences (.989 vs .978, $t(99) = 49$, $p < 10^{-15}$).[11]

2.5 Interim summary

van Tiel (2013) argued that the data in Chemla and Spector (2011) deserved a better analysis, one quantifying properly not only ambiguity but also typicality. He furthermore proposed a specific model of typicality and typicality projection and provided the means to quantify the role of typicality so construed. We provided a complete analysis of the data. Our analysis is conservative in the sense that we assumed that typicality as construed by van Tiel is independent of scalar implicatures. But in our analysis we explicitly *evaluated* the role of the readings predicted by localist theories. Given that typicality à la van Tiel explains a great deal of the data, taking it into account puts us in a difficult situation to demonstrate that some other factor also plays a role. Yet, the results demonstrate that local scalar implicatures have an explanatory role, beyond typicality.

3 Moving forward: a new method to evaluate ambiguities

We submit that the method offered by van Tiel (2013) can be used more generally to test cases of ambiguity in a systematic manner and assess the relative accessibility of the different

---

[9]Note for instance that taking into account only the literal reading *decreases* the fit, compared to the model where no reading is taken into account, i.e. van Tiel's own simulation. This holds both for the *some* sentences (.994 vs .989, $t(99) = 17$, $p < 10^{-15}$) and for the *or* sentences (.994 vs .978, $t(99) = 72$, $p < 10^{-15}$).

[10]We report statistics from paired t-tests (see also footnote 8), but other statistical tests lead to the same conclusions.

[11]In fact $\mathcal{L}$ is a better predictor than any other set of readings (including the empty set, which corresponds to van Tiel's original simulation) for EVERY-SOME sentences and it is outperformed only by the set of readings which excludes the literal reading ({global, literal}) for EVERY-OR sentences. See also Table 1 below.

readings. We show how this applies to the case of scalar implicatures (and further buttresses our conclusion). Instead of assigning a constant reward for each reading, we can ask what weights should be given to each reading in order to obtain the best possible fit. If the local reading plays no role in the data, it should not receive a significant positive weight. So we ran a second analysis, still based on van Tiel's elegant method and factoring out the contribution of typicality (as construed by van Tiel), to evaluate the relative weights of the different readings. We proceeded as follows.

(11)   Simulation 2: evaluating the relative salience of different readings
   a.   Choose a triplet of ordered values for $\rho(\text{none})$, $\rho(\text{all})$, $\rho(\text{some but not all})$, complying with Assumption 2.
   b.   Compute the typicality value ρ_E for the "every" sentence in the seven conditions of Chemla and Spector's data, given the triplet chosen in (a), and the rule in Assumption 1.
   c.   Find out what's the best way to fit C&S's data by a linear model using as predictors: the typicality value obtained from (b) and the truth-value of each of the different readings. We thus obtain the best possible 'weights' α_{typ}, α_{lit}, α_{glob}, α_{loc} for respectively typicality, the literal reading, the global reading and the local reading. Formally, the model is as follows:

$$
\begin{aligned}
\text{C\&S's score (cond)} = \quad & \alpha_{\text{typ}} \times \rho_E(\text{cond}) \\
+ \; & \alpha_{\text{lit}} \; \times \text{TV}_{\text{lit}}(\text{cond}) \\
+ \; & \alpha_{\text{glob}} \times \text{TV}_{\text{glob}}(\text{cond}) \\
+ \; & \alpha_{\text{loc}} \; \times \text{TV}_{\text{loc}}(\text{cond})
\end{aligned}
$$

   in which "$\text{TV}_x(\text{cond})$" is the truth-value of reading 'x' in condition 'cond'.
   d.   Do this 100 times to obtain an estimation (mean and distribution) for each of the different weights: the contribution of typicality and the contribution of each of the readings.

Again, we also ran the same simulation for the EVERY-OR case. We obtain an estimate of the weights that each (putative) reading should receive in order to best match the data. If a reading R is not useful, the average weight $\bar{\alpha}_R$ associated with it should not be different from zero. Note that, by adding predictors to the model, we are guaranteed to improve the fit compared to van Tiel's initial model (contrary to the analysis presented in 2.4). The critical question, however, does *not* have to do with the fit of the model (the r correlation cœfficient), but is rather whether the weight assigned to each reading is positive. Mathematically speaking, this is in no way a guaranteed outcome. Even though adding predictors to a model can only improve the fit, nothing guarantees in advance that the weights of different predictors will be positive rather than negative. Finding a significant positive weight for a reading would thus provide evidence for the role played by this reading.

Crucially, the local reading was given a statistically significantly positive weight, showing that it exists, both in the case of EVERY-SOME sentences and EVERY-OR sentences ($\bar{\alpha}_{\text{loc}} = 4.97$ and 3.31, respectively, $ts > 11$, $ps < 10^{-20}$). Under this analysis, the local reading is in fact given the heights weight (all $ts > 10$ in the relevant pairwise comparisons). The full results are reported in Table 1.We conclude that if van Tiel's typicality measure is the correct one, then the local reading plays a role in explaining our data on top of typicality.

	'some'		'or'	
Literal reading	$\alpha_{\text{lit}} = 22,$	$t(99) = 170$	$\alpha_{\text{lit}} = 14,$	$t(99) = 110$
Global reading	$\alpha_{\text{glob}} = 20,$	$t(99) = 550$	$\alpha_{\text{glob}} = 15,$	$t(99) = 420$
Local reading	$\alpha_{\text{loc}} = 24,$	$t(99) = 240$	$\alpha_{\text{loc}} = 25,$	$t(99) = 260$

Table 1: Weights obtained from simulation 2, as described in (11). All p-values are below 10^{-100}.

The combination of Chemla & Spector's methodology with van Tiel's method and the present modification of it could offer a way to evaluate the relative salience of different readings quite generally. Concerning scalar implicatures, it may help address an aspect of the debate which has often been oversimplified: one can often read (e.g., in Geurts and Pouscoulous, 2009) that the local reading should be the only reading according to localist approaches. This is in fact not so (as we had stressed in Chemla and Spector, 2011). But, in any case, no explicit account (from no camp) exists as to which readings are preferred and, practically, the relevant data would be difficult to assess without a method such as the one we describe above.

4 Revisiting van Tiel's data, towards a linguistic theory of typicality

We have shown that, *even* if van Tiel's views on typicality were entirely correct, one could demonstrate that scalar implicatures play a role in the task our subjects participated in. Now we would like to come back to the question as to whether van Tiel's typicality measure is in fact the right one. In connection to the original debate, the point is this: we cannot assume that the experimental data on the basis of which van Tiel constructed his typicality measure does not already include effects of the truth-value of various readings (including strengthened readings). More generally, we will argue that a conceptually motivated theory of typicality can be developed.

In section 4.1, we will have another look at the data concerning what van Tiel calls the 'typicality structure' of 'some.' By (loose) analogy, one could relate this issue with what has been called the 'triggering problem' in the realm of presuppositions. We will propose a simple model that relates the typicality value for a sentence in a given situation to the distance between the given situation and true/false situations. Such a notion of distance is natural in the realm of 'typicality', and we will show that the simple resulting model provides a good account of van Tiel's data.

In section 4.2, we will move to the question of typicality projection (under 'every'). We will proceed in two steps. First, we will show that there is no empirical reason to characterize typicality projection under universal quantification by means of the harmonic mean, as opposed, say, to some other type of mean (section 4.2.1). This will lead us to pursue a more principled approach of the study of typicality projection and, again, propose a simple model for it (section 4.2.2).

Finally, in section 4.3, we will show that the two models (for the triggering problem and the projection problem, so to speak) are in fact based on the same primitive assumptions. We will thus propose a general version of this model and tentatively submit it as a proposition for a general linguistic theory of typicality.

4.1 Typicality structure of 'some'

Let us first look at the typicality structure of 'some'. van Tiel reports the results of a task where subjects had to rate the sentence 'some circles are black' when paired with different pictures containing 10 circles, and where the number of black circles varied. The mean scores, depending on the number of black circles, are given in Fig. 3.
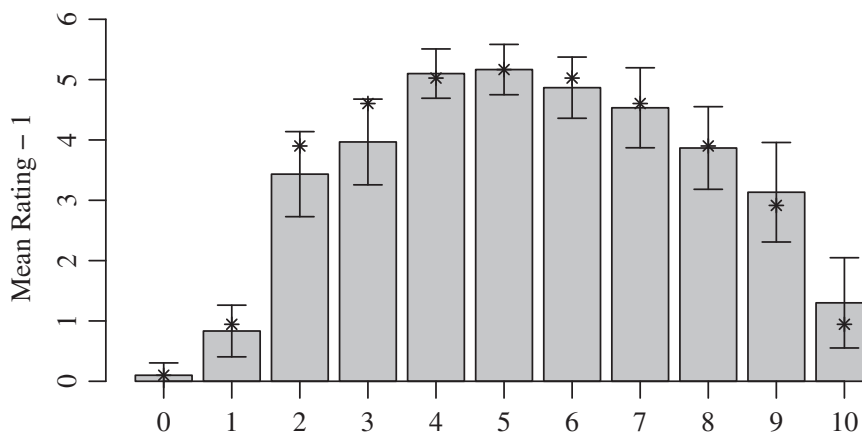

Figure 3: Figure 6 from van Tiel (2013). *Mean typicality rating for the sentence 'Some of the circles are black' in situations with ten circles. (...)*

On this basis, van Tiel assumed that, in our own task, strong verifiers (i.e. cells in which the letter is connected to some but not all of its circles) are more typical instances of the predicate 'be connected with some of its circles' than weak verifiers (i.e. cells in which the letter is connected to all of its circles), and that weak verifiers are more typical instances of this predicate than cells in which the letter is connected to no circle. This is of course consistent with van Tiel's data about 'some', but the scores obtained by van Tiel for 'some circles are black' may in fact simply reflect the conjunction of the literal reading of the sentence and the 'not-all' scalar implicature. If this were so, then van Tiel's simulation of our results for EVERY-SOME sentences could be plausibly argued to integrate the potential effect of strengthened readings, including even the 'local' reading. In this case, not only would it be impossible to conclude that typicality is sufficient to explain our data (something we have excluded above), it also could not be claimed that typicality considerations provide a useful explanation for Chemla and Spector's data that does not resort to the 'local' reading of EVERY-SOME sentences to begin with.

It is thus crucial even to a weaker version of van Tiel's claim, namely that typicality as an independent factor matters, that the results he gathered for 'some' are independent of scalar implicature computation. As an argument for the view that the 'not-all' scalar implicature plays no role in the 'typicality structure' of *some*, van Tiel notes that cases where about half of the circles are black are judged 'more typical' than cases where, say, 3 or 8 out of the 10 circles are black, even though in all these cases the 'some-but-not-all' reading of the sentence is true. But this is clearly a *non sequitur*. At best, van Tiel's data for *some* show that the strengthened, 'some but not all'-reading does not explain all of the data, not that it plays no role. In fact, based on this spurious logic, van Tiel should also have concluded that the *literal* reading plays no role in the typicality structure of *some* (since the literal reading cannot either explain the

different ratings received by pictures with 5 black circles vs. those with 3 or 8 black circles). Yet, strikingly, van Tiel's model of typicality for 'some circles are black' does in fact include the truth-value of the literal reading of the sentence as a factor. Indeed, van Tiel offers the following model to capture the typicality structure of 'some':

$$\rho_{\text{SOME A B}} = 1 - \frac{1}{Z}[(S - 5)^2 + v_S(\text{`Some A B'})],$$

where Z is a normalizing factor ensuring that the final value will lie between 0 and 1, S is the number of black circles, and v_S('Some A B') is the **truth-value of the literal reading**.

The view is that a situation with 5 black circles is 'prototypical', so that $(S - 5)$ can be viewed as the distance between the picture and the 'prototypical' situation. Testing this model against his experimental data, van Tiel obtains an excellent correlation $r = .95, p < .01$. The fact that prototypical situations are those that are more 'central' is thus taken as a primitive in van Tiel's system. Ideally, however, it would be preferable to derive this 'central' tendency from an independent property of the word 'some', or possibly some more general cognitive bias. To us, the results reported by van Tiel seem to be expected if speaker's judgments for each picture are entirely driven by a) the strengthened reading ('some but not all') and b) subjects' perception of the 'distance' between a situation and minimally different situations making this reading true or false. More specifically, here is a simple set of assumptions that can capture van Tiel's results:

1. The sentence 'Some circles are black' is most saliently understood as equivalent to 'At least two circles are black and not all circles are black'. (This assumption will be relaxed later on in section 4.3 when we take into account possible ambiguities).

2. In van Tiel's graded judgment task, the subjects' rating of a picture is a linear function of a) the 'distance' to the 'closest' case making the sentence false (one expects a positive coefficient here), b) the distance to the closest case making the sentence true (one expects a negative coefficient here), c) the truth-value, relative to the picture, of the proposition under consideration (again, a positive coefficient is expected).

3. The notion of 'distance', for the kind of pictures used in van Tiel's experiments, can be plausibly approximated as follows: given a certain picture, the distance to the 'closest' case making the sentence true (resp. false) is the smallest number n (possibly 0) such that changing the color of n circles in the picture would make the sentence true (resp. false).[12]

In other words, we suggest the following linear model:

(12) $\quad \rho_{\text{`some circles are black'}}(\text{picture}) = \quad constant$
$\qquad\qquad + \quad \alpha \times \text{Truth-value}$
$\qquad\qquad + \quad \beta \times \text{Distance to the closest true case}$
$\qquad\qquad + \quad \gamma \times \text{Distance to the closest false case}$

With such a model, on the basis of the strengthened reading ('at least two circles are black, and not all are'), we obtain an excellent fit of van Tiel's results reported in Fig. 3 ($r > 99, p <$

---

[12]We should note however, that what we know of numerical cognition (e.g., Dehaene, 1997) leads us to expect that the perceived distance between 9 and 10 is in fact smaller than that between 1 and 2.

10^{-5}). The weight of the predictor 'Truth-value' is, as expected, positive, but is not significantly different from 0 ($\alpha = 0.46, p = .24$) and the weights for 'distance to the closest true case' and 'distance to the closest false case' both have the expected sign and are significantly different from 0 ($\beta = -0.65, p < .05, \gamma = 0.42 : p < .0005$). Hence, van Tiel's typicality data can be explained by what we see as minimal assumptions about the role of the distance between the current situation and true and false situations.

4.2   A model of van Tiel's data regarding projection under 'every'.

We now turn to the projection of typicality under 'every'. First, we will show that there is some arbitrariness in van Tiel's proposal (section 4.2.1). Second, we will propose an alternative model, whereby in fact typicality projection is based on the same principles as the derivation of the typicality structure of *some* (section 4.2.2).

4.2.1   'Typicality' projection under 'every': why the harmonic mean?

van Tiel's rule for typicality projection under 'every' is based on the harmonic mean. The reason why he chose the harmonic mean rather than some other type of mean was because it provided a good fit (with a specific choice of parameters, $r = .97$) with respect to the data he gathered concerning the projection of typicality under 'every'. Note however that it did not provide the best possible fit among a set of reasonable candidates: the geometric mean provides an even better fit, $r = .98$.[13] We suspect that the reason why van Tiel used the harmonic mean is that it provided a better fit than the geometric mean in his simulation of our own data, as we will see shortly. However, we now have provided evidence that even if we accept that van Tiel's own model of typicality projection under *every*, the various readings hypothesized by localists play a role in explaining the data. So the fact the harmonic mean, in the context of van Tiel's simulation, led to a better fit than the geometric mean for modeling our original data gives no strong reason to assume that typicality projection under *every* should be modeled in terms of the harmonic mean. Rather, what we need to assess is which type of mean is the best *when readings are taken into account in the simulation*.

We ran various analyses to compare different choices of means. First, we ran simulations *à la* van Tiel, using the harmonic mean, but also the arithmetic mean and the geometric mean. Second, we ran simulations in which the various hypothesized readings are taken into account. We took the readings in consideration as in section 3, by choosing the best possible weights for the readings each time. Contrary to what we did in section 3, we were now interested in the quality of the fit of the resulting model, when obtained using three different types of means for the computation of the typicality value: the harmonic mean (as before), the arithmetic mean or the geometric mean.

The results are not robust. If we do not take into account the readings (but why wouldn't we?), the harmonic mean seems to provide a better fit for EVERY-SOME sentences (Table 2). Hence, it certainly was a good choice from van Tiel's perspective. But this result does not generalize. First, it does not generalize to EVERY-OR sentences, for which the geometric mean provides a better fit ($t(99) = -3.8, p < 10-3$). Second, if we take the readings into account (Table 3), the harmonic mean is the one that provides the poorest (although still very high) fit from the three types of means with both EVERY-OR and EVERY-SOME sentences (all $ts > 50$).

---

[13]We recovered van Tiel's data from the graphs presented in the paper, i.e. by measuring bar lengths by hand, with a ruler.

	EVERY-SOME	EVERY-OR
Harmonic mean	.991	.988
Geometric mean	.970	.994
Arithmetic mean	.815	.857

Table 2: Correlation cœfficients for models based on different types of means, without taking possible readings into account (à la van Tiel).

	EVERY-SOME	EVERY-OR
Harmonic mean	.9911	.9837
Geometric mean	.9966	.9931
Arithmetic mean	.9942	.9882

Table 3: Correlation cœfficients for models based on different types of means, taking possible readings {literal, global, local} into account.

Given such results, the data discussed here provide no reason to pick the harmonic mean over the geometric mean. In fact, it is not clear that a mean-based theory of typicality projection under 'every' is adequate and can be motivated by empirical arguments of the kind we try to produce above or by independent conceptual reasons. In fact, it is hard to extend a mean-based theory of projection to other quantificational or embedding environments. In the next subsection, we propose an alternative model of the typicality projection under 'every', which is conceptually simple and which is principled, in the sense that it can be generalized to any environments without specific assumptions about these environments (This desiderata mimics a vivid discussion about the explanatory power of various accounts of presupposition projection, see, e.g., Rothschild, 2008 and references therein).

4.2.2 Typicality projection under 'every': a model

Let us take a closer look at van Tiel's results about typicality projection under *every*. van Tiel asked for 'typicality judgements' for sentence-picture pairs with the sentence 'Every circle is black', and where the picture contained a certain number of black circles and a certain number of white circles (out of ten circles). These results are summed up by the graph reproduced in Fig. 4.

van Tiel models these data in the following way: for every picture, a) he assigns a very low typicality value (e.g., .1) to each white circle, and a very high typicality value to each black circle (e.g., .95), and b) he takes the harmonic mean (as we have already discussed) of the typicality values assigned to each individual circle to be the typicality value of the picture relative to the universally quantified sentence. As we saw, the reason why the harmonic mean is used is post-hoc: it is better than the arithmetic mean (but not the geometric mean) in predicting the striking 'leap' between the last two bars (i.e. the fact that the score of a picture increases much more when you go from 9 black circles to 10 black circles than when you go, say, from 4 black circles to 5 black circles).[14]

---

[14]The precise prediction of van Tiel's model based on the harmonic mean depends on the typicality value associated to each individual circle relative to the predicate 'black'. van Tiel reports that "the correlation exceeds
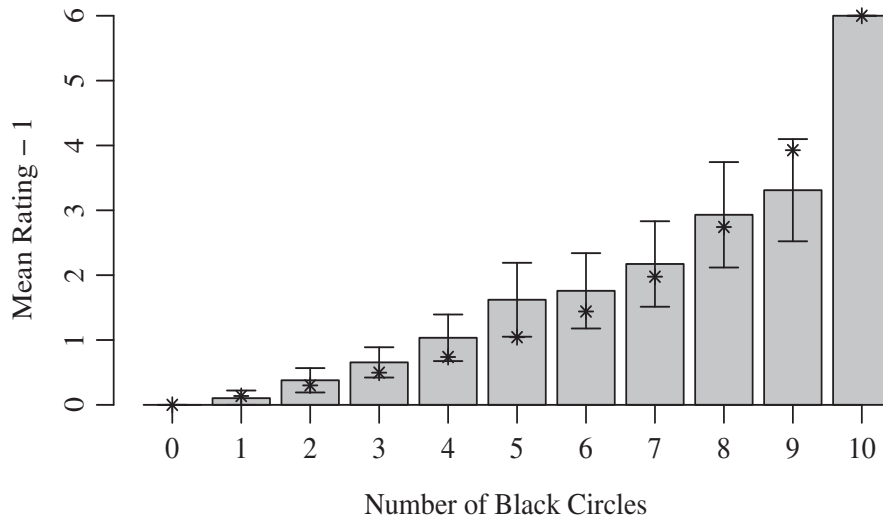
Figure 4: Figure 5 from van Tiel (2013). *Mean typicality rating for the sentence 'Every circle is black' in situations with ten circles. (...)*

To us, it seems that van Tiel's data can be efficiently described in two steps: on the one hand, in all the cases where the sentence is false, the rating of a given picture increases linearly with the number of black dots, and, on the other hand, in the only case where the sentence is true, there is some kind of an added bonus. That is, we may propose a very simple model in which the rating of a picture relative to the universal sentence is a linear function of a) the number of black dots in the picture, and b) the truth-value of the sentence relative to the picture (and we expect positive weights for these two factors):

(13)
$$\rho_{\text{'some circles are black'}}(\text{picture}) = \begin{aligned} & constant \\ + \ & \alpha \times \text{Truth-value} \\ + \ & \beta \times \text{Number of black dots} \end{aligned}$$

Since the number of black circles in a picture can be viewed as a measure of how close the picture is to a 'true case' (the more black dots, the closer are we to a case making true 'Every circle is black'), this interpretation amounts to saying that the observed ratings correspond to the combination of a categorical judgment (true vs. false) and a 'typicality' judgment that is itself driven by truth-conditional considerations. In fact, with such a linear model, on the basis of van Tiel's data, we obtain an excellent fit ($r > .99, p < 10^{-7}$), with both predictors receiving a positive weight significantly different from 0 ($\alpha = 1.74, \beta = .26, ts > 9, ps < 10^{-5}$).

4.3 Discussion: towards a general model of typicality

4.3.1 A single model for 'every' and 'some'

In contrast with van Tiel who proposed two unrelated (and unmotivated) models for 'every' and for 'some', we in fact used exactly the same model in both cases. Although it is possibly not visible from the comparison of the equations in (12) and (13), it is so for two reasons. First,

---

r=.93 for almost all possible values such that $\rho(\text{black}) > \rho(\text{white})$." With $\rho(\text{black}) = .95$ and $\rho(\text{white}) = .1$, he obtains $r = .97$.

'Number of black dots' in the case of 'every' is equal to 10-'distance to a false case', so that 'Number of black' in the model for 'every' plays the same role as 'distance to a false case' in the model for 'some'. Second, 'distance to the closest false case' turns out to be identical to 'truth-value' for 'every' (it is so because all cases but one are false so that the distance is either 0 if the sentence is false or 1, if the situation happens to be the true situation). For this reason there is no need to include an independent 'distance to a false case' in the model for 'every'.

So we can now evaluate how our model is able to predict the data for both 'some' and 'every' together. Putting together van Tiel's data for both 'some' and 'every', we obtained 22 data-points (2 sentences × 11 pictures, each picture containing from 0 to 10 black circles). We ran a linear regression based on our model. We added as an additional predictor the sentence-type ('some' vs. 'every'). This amounts to allowing for different acceptation biases for different kinds of sentences, which could also be captured by allowing for the constant in the following equation to depend on the sentence. See also footnote 15). Concretely, we tested the following model:

$$
\begin{aligned}
(14) \quad \rho(\mathrm{S}, \mathrm{w}) = \quad & constant \\
+ \quad & \alpha \times \text{Truth-value of S in w} \\
+ \quad & \beta \times \text{Distance between w and the closest true case} \\
+ \quad & \gamma \times \text{Distance between w and the closest false case} \\
+ \quad & \delta \times [\text{S='some of the dots are black' vs. S='every circle is black'}]
\end{aligned}
$$

We obtain again an excellent fit ($r > .99, p < 10^{-14}$), and each predictor is assigned a weight significantly different from 0 (all $ps < 10^{-5}$), and of the expected sign ($\alpha = 1.10, \beta = -.27, \delta = .39$).[15]

In essence, these results suggest that there might be a very simple general model of typicality. This model makes typicality structure dependent on the truth-conditions of the relevant sentences and on some notion of distance between situations. Ideally, we would like to apply this model to our test cases. There will be several complications to do so, one of them is the fact that this model does not yet take ambiguity account.

4.3.2 A model of typicality, taking ambiguity into account

In this section, we want to take into account possible ambiguities. In essence, we are going to assume that if a sentence has several possible interpretations, their typicality values are going to be combined according to the relative saliency of these different readings. So, consider a sentence S with two possible readings: $R1$ and $R2$. Assume that we can define saliency weights for these readings w_1 and w_2, in a way that $R1$ is more salient than $R2$ to the same extent as w_1 is larger than w_2. Thus the typicality value of S, we assume, can be captured as the weighted sum of the typicality values of the readings:

(15)

---

[15]Additionally, we obtain $\gamma = 1.69, t = 14$, which basically shows that 'every' sentences received an overall higher rating.

$$\begin{aligned}
\rho_{\mathrm{S}}(\mathrm{w}) = \ & constant \\
& + \ w_1 \times (\quad \alpha \times \text{Truth-value of } R1 \text{ in w} \\
& \qquad\qquad + \beta \times \text{Distance between w and the closest true case for } R1 \\
& \qquad\qquad + \gamma \times \text{Distance between w and the closest false case for } R1) \\
& + \ w_2 \times (\quad \alpha \times \text{Truth-value of } R2 \text{ in w} \\
& \qquad\qquad + \beta \times \text{Distance between w and the closest true case for } R2 \\
& \qquad\qquad + \gamma \times \text{Distance between w and the closest false case for } R2)
\end{aligned}$$

In this model, the notion of typicality is captured by the parameters α, β and γ. In this model, this triplet of values is the essence of how typicality emerges from truth-value and distance to true/false cases. These parameters are here assumed to be the same for all possible readings of a sentence. Similarly, our joint model for 'some' and 'every' above was such that the triplet (α, β, γ) was assumed to be the same for both kinds of sentences.

Let us go back to the case of 'some'. 'Some circles are black' has two possible interpretations a weak (literal) interpretation and a strengthened ('not all') interpretation. In section 4.1, we proceeded as if it had only the strong interpretation, but we can now take the two interpretations into account. Applying the non-linear model , the estimates for the weights of the strong and weak interpretations are $w_{\mathrm{strong}} = .086$ and $w_{\mathrm{weak}} = .00041$. As a result, the strong interpretation is 200 times more salient than the weak interpretation, or less controversially, it will not play much of a role in the model. As a result, it is not a surprise that the result of this new model is not different from the results from the simpler model. In particular, the (α, β, γ) triplet we obtain is (5.6, -7.4, 4.9), which is roughly equal to w_{strong} times the triplet obtained in section 4.1 (.46, -.65, .42) (the multiplication gives: 5.4, -7.6, 4.9).

4.3.3 Applying the model to our test cases?

We now have a general model of typicality, as summarized in (15). For any sentence, however complex, the way to solve the typicality projection problem would involve a) determining the various readings of the sentence, and b) defining a metric thanks to which the notion of 'distance to a true/false case' can be defined.

We already suggested in our initial paper that the rating of a given picture for EVERY-SOME sentences should depend on the distance of the picture to situations making various readings true or false, and we should now have the means to objectively evaluate this claim. However, we cannot immediately test how such a model fares when applied to the type of data we discussed in our original paper. There are two technical problems:

1. The relevant notion of 'distance' is much harder to define in the case of complex pictures of the type we used than it is for the kind of pictures tested by van Tiel. Note for instance that, on the one hand, a 'strong verifier' (a cell in which the letter is connected with some but not all of its circles) is intuitively closer to a falsifier (a cell in which the letter is connected to no circle) than a weak verifier is. So a picture that contains only strong quantifiers is in some sense quite close to a case of falsity. But, on the other hand, such a picture is one where all of the three readings we hypothesized are true. There is no straightforward notion of distance between pictures that we could confidently use in this case without further providing independent arguments (see also Cummins, 2014, which phrases this difficulty in an elegant way, making it a 2D problem.).

2. If we want to test the role played by the various readings, the relevant predictors in the linear regression will include: a constant, the weights of the different readings and the (α, β, γ) triplet of factors, i.e. in principle 7 different predictors for only 7 data-points if we consider only EVERY-SOME sentences.

Hence, we do not believe that the available set of data allows us to test our general model of typicality in a serious way. But more data can be collected to (i) ground a realistic notion of distance and (b) test our corresponding general model of typicality.

One advantage of our model is that it applies to virtually any sentence. The mean-based account was oriented towards the 'every' sentences. Our model is not. One appreciable consequence of it is that it could make predictions for EXACTLY ONE-SOME sentences of our initial experiments, which we argued came at the core of the scalar implicature issue. We see a broader application of the current typicality model, including application to these non-monotonic cases and others, as an interesting direction for future research.

5 Summary and conclusion

In this paper, we have made use of methodological and conceptual insights of van Tiel (2013). First, we have re-assessed evidence put forward in favor of the existence of local scalar implicatures. We conclude that the evidence is solid. We would like to note that we have left aside other kinds of evidence along the way. Most prominently, we did not discuss non-monotonic environments, which we argued in Chemla and Spector (2011) provide a much better test-case. We also did not discuss other arguments coming from the interaction between scalar implicatures and Hurford's constraint (see Chierchia et al., 2012; Singh, 2007; Chemla, 2013) or from intermediate scalar implicatures (see Sauerland, 2012).

Second, we have shown that the results obtained by van Tiel regarding the typicality structure of *some* and *every* can be accounted for by a very simple model in which the rating of a sentence relative to a picture is an increasing linear function of its truth-value, how close the picture is to a case making the sentence true, and how far the picture is from a case making the sentence false. This suggests to us that our initial interpretation of our own data was on the right track.

Third, we have embraced what we believe is an interesting issue raised by van Tiel's project, quite indepenently of scalar implicatures, namely the issue of typicality projection. Presupposition projection and the interactions of scalar items with various linguistic operators have been studied extensively. There is no reason why typicality projection could not follow the same path.

References

Chemla, E. (2008). *Présuppositions et implicatures scalaires: études formelles et expérimentales*. Ph. D. thesis, ENS.

Chemla, E. (2009a). Presuppositions of quantified sentences: experimental data. *Natural Language Semantics 17*(4), 299–340.

Chemla, E. (2009b). Similarity: towards a unified account of scalar implicatures, free choice permission and presupposition projection. Under revision for *Semantics and Pragmatics*.

Chemla, E. (2009c). Universal implicatures and free choice effects: Experimental data. *Semantics and Pragmatics 2*(2), 1–33.

Chemla, E. (2013). Apparent hurford constraint obviations are based on scalar implicatures: An argument based on frequency counts. Ms. CNRS, ENS, LSCP Paris.

Chemla, E. and B. Spector (2011). Experimental evidence for embedded implicatures. *Journal of Semantics 28*(3), 359–400.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In A. Belletti (Ed.), *Structures and Beyond*. Oxford University Press.

Chierchia, G. (2006). Broaden Your Views: Implicatures of Domain Widening and the 'Logicality' of Language. *Linguistic Inquiry 37*(4), 535–590.

Chierchia, G., D. Fox, and B. Spector (2012). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics.

Clifton, C. J. and C. Dube (2010, July). Embedded implicatures observed: A comment on Geurts and Pouscoulous (2009). *Semantics and Pragmatics 3*(7), 1–13.

Cummins, C. (2014). Typicality made familiar: A commentary on Geurts and van Tiel (2013). *Semantics and Pragmatics*. In press.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. Oxford University Press.

Fox, D. (2007). Free Choice and the theory of Scalar Implicatures. In U. Sauerland and P. Stateva (Eds.), *Presupposition and Implicature in Compositional Semantics*, pp. 537–586. New York, Palgrave Macmillan.

Geurts, B. and N. Pouscoulous (2009). Embedded implicatures?!? *Semantics and Pragmatics 2*(4), 1–34.

Landman, F. (1998). Plurals and Maximalization. In S. Rothstein (Ed.), *Events and Grammar*, pp. 237–271. Kluwer, Dordrecht.

van Rooij, R. and K. Schulz (2004). Exhaustive Interpretation of Complex Sentences. *Journal of Logic, Language and Information 13*(4), 491–519.

Rothschild, D. (2008). Making dynamics semantics explanatory. Ms. Columbia University.

Sauerland, U. (2012). The computation of scalar implicatures: Pragmatic, lexical or grammatical? *Language and Linguistics Compass 6*(1), 36–49.

Singh, R. (2007). On the Interpretation of Disjunction: Asymmetric, Incremental, and Eager for Inconsistency. Ms., MIT.

Spector, B. (2003). Scalar implicatures: Exhaustivity and Gricean reasoning. In B. ten Cate (Ed.), *Proceedings of the Eigth ESSLLI Student Session*, Vienna, Austria. Revised version in *Questions in Dynamic Semantics*, eds. M. Aloni, P. Dekker & A. Butler, Elsevier, 2007.

Spector, B. (2006). *Aspects de la pragmatique des opérateurs logiques*. Ph. D. thesis, Université Paris 7.

van Tiel, B. (2013). Embedded scalars and typicality. *Journal of Semantics*. In press.