Maximality and Definite Plurals - Experimental Evidence

Florian Schwarz - University of Pennsylvania

Abstract. Statements with plural definite descriptions are often assumed to come with a maximal interpretation, requiring all the individuals meeting the description to have the property expressed by the predicate in the sentence. However, there clearly is some variability with respect to this requirement, as non-maximal interpretations seem to be possible in various circumstances as well. This paper presents experimental evidence informing the relationship between maximal and non-maximal interpretations of definite plurals. I suggest that the results are best captured by a view that sees the semantics of plural definites as involving maximality, but which allows for non-maximal interpretations, even with distributive predicates, by some type of pragmatic mechanism. Various possible choices of such mechanisms are discussed, though the question of which one is best suited to account for the data is not conclusively resolved. However, the advantages over other accounts, which assume a non-maximal semantics combined with pragmatic strengthening when needed to derive maximal readings, with respect to the present results seem rather clear.

Keywords: definites. plurals. maximality. experimental semantics. reaction times.

1. Introduction

The semantics of plural definite descriptions, such as *the boys*, has played an important role in a number of theoretical debates, including ones concerning the types of things involved in natural language meanings. In particular, theories in this domain have to encompass a view on what sorts of plural entities there are and how they relate to verb meanings. This, in turn, leads to distinctions between distributive and collective interpretations, as well as potential lexical distinctions between classes of predicates and a host of pragmatic factors. While there is an extensive literature on these and other intricate theoretical issues, with substantial progress over the years and sophisticated technical proposals to deal with a wide range of relevant data, some of the basic empirical properties of plural definites have not been studied systematically in great detail.

One characteristic property of plural definites relates to the notion of maximality. Many theoretical proposals assume that definite plurals receive a maximal interpretation. For example, *The boys left.* is typically seen as expressing that all the (contextually relevant) boys left. At the same time, however, it also seems rather clear that there is a fairly wide range of cases where maximality does not seem to hold. For example, the sentence *The boys are building a raft.* does not seem to require that all of the boys are involved in the building. It's possible that a few of them are doing something else. While many of the accounts in the literature do make reference to at least some issues relating to the apparent variation between maximal and non-maximal readings, it is not clear that we have a comprehensive understanding of the range of factors affecting maximality, let alone a unified understanding of anything close to the full range of the relevant data.

On a general level, one way to frame a discussion of maximality, based on the assumption that both maximal and non-maximal judgments are attested, is to ask what the status of the two types of interpretations is and how they relate to one another. Logically speaking, there are the following possibilities: they could be entirely independent from one another, and both reflect separate semantic entries for (the plural version of) the definite article *the*, i.e., there could be a genuine ambiguity. Secondly, the dimension of maximality could be in some sense underspecified in the semantics, and be decided (or left underspecified when no decision is necessary) in any given case based on a variety of factors (which could be both linguistic and non-linguistic). Finally, one of the readings might be a basic one, with the other derived from it in some form or other, most likely in pragmatic terms. I take it that the first option, that of a genuine ambiguity, has very little appeal, as it does not offer any explanatory strategy for understanding the distribution of the different readings. The other options seem at least in principle viable, and have been pursued in the literature.

The present paper aims to contribute to the discussion of which of these views is appropriate by presenting empirical evidence relevant for understanding the relationship between maximal and non-maximal interpretations. The evidence stems from a series of experiments that involved truth-value judgments (TVJ) for sentences in the context of simple visual displays consisting of colored shapes. For example, subjects would see a mix of black and gray circles and then be asked to evaluate the truth of a sentence such as *The circles were black*. The details of the design will be laid out below. Before delving into those, I will briefly review some of the major proposals for analyzing the semantics of plural definites, and how they relate to maximal and non-maximal interpretations respectively. But to foreshadow the general thrust of the line of argument pursued based on the experimental results, we will see that a), non-maximal interpretations are to some extent available even with basic distributive predicates such as color adjectives, b), a variety of pragmatic factors modulate this availability, and c), maximal responses are generally faster than non-maximal ones. I will argue that, taken together, these results speak in favor of an account that sees maximality as part of the lexically encoded semantics, and which derives (at least one kind of) non-maximal interpretations by means of a process of pragmatic weakening.

2. Background

2.1. 'Strong' Accounts of Plural Definites

The starting point for most accounts of plural definites is the semantics for singular definites. One pre-dominant family of theories sees these as involving uniqueness; the family line following Strawson (1950) assumes that uniqueness is presupposed, rather than asserted, rendering the following type of meaning for the definite article, which makes it denote in the domain of entities:[1]

[1]The extension of uniqueness to maximality to be outlined here can also be implemented in a Russellian framework, where uniqueness is part of the asserted content (Neale, 1990), though this runs into some issues with collective readings.

(1) a. $[\![\textit{the}_{Sg}]\!] = \lambda P_{\langle e,t\rangle}.\iota x.P(x)$ b. $[\![\textit{the}_{Sg}\ \textit{boy}]\!] = \iota x.\text{boy}(x)$

On a presuppositional view, this entry comes with a requirement that there is only one (contextually relevant) individual of whom the predicate P / $[\![\textit{boy}]\!]$ holds. Based on work on plurality by Sharvy (1980) and Link (1983), the domain of entities can be enriched so as to include plural individuals, which count as entities of their own right, but have multiple atomic individuals as their parts. These can be seen as sum individuals or sets, which allow access to their part structure. Groups, on the other hand, are another type of entity consisting of multiple atomic individuals that have been considered in the literature (perhaps most prominently Landman, 1989). Their hallmark property is that they do not allow access to their parts, the idea being that there is more to the whole group than the individuals that are part of it (e.g., a team seems to have an existence of its own that remains constant, even if team members change over time).

Now, a natural extension of uniqueness-based accounts of singular definites to plural definites, pioneered by Sharvy and Link, is to re-frame uniqueness in terms of maximality: rather than having the definite article map a singleton set to the atomic member of that set, it can now be seen more generally as mapping any set, which may include plural individuals, onto the maximal individual in that set (the notion of maximality can be formally defined in terms of the part-relation, the maximal individual being the one that all other individuals in the set are part of). This is commonly expressed with the σ operator:

(2) $[\![\textit{the}]\!] = \lambda P.\sigma x.P(x)$

Maximality thus provides a unified view on singular and plural definites. Singular definites will combine with singular predicates, which only consist of atoms. For there to be a maximum individual in a set of atoms, there can only be one individual, so we capture the equivalent of the uniqueness account above. In cases where the predicate is plural and contains plural individuals, the definite maps that set onto the maximal plural individual contained in it (if there is one).

Given the notion of 'maximal plural individual', it may already appear clear that accounts along these lines assign a maximal semantics to plural definites. However, to fully flesh out the story, something more needs to be said about how verbal predicates combine with plural noun phrases. While there are again various options, one commonly taken path is to say that there is a distributive operator, ^{D}, which can attach to verb phrases, and which introduces universal quantification over parts or subsets, depending on your view on what plural individuals are (see Link, 1983; Landman, 1989, 1996; Lasersohn, 1995; Schwarzschild, 1996, for several different versions). For purposes of exposition, I follow the format of Brisson (1998):

(3) $[\![^{D}]\!] = \lambda P.\lambda x.\forall x[y \in x \rightarrow P(y)]$

Combined with a verb, this will render a function that will yield true for any (singular or plural) subject argument iff it yields true for all the elements (or parts) of the denotation of that argument. Now we have a strictly speaking maximal semantics, as on this analysis,

(4) *The boys* ^{D}*left.*

requires that every boy left. This is the so-called *distributive* reading of the sentence. However, note that the D operator is not generally speaking required. If we leave it out, we get the so-called *collective* interpretation, where the predicate is only required to hold of the plural individual as a whole. This accounts for cases like

(5) *The boys built a raft.*

For a collective interpretation to be true, it is not only commonly argued that the predicate need not hold of every individual separately, but also that not every individual that is part of the relevant plural individual has to have any actual involvement with the activity in question. This, then, provides one way of allowing non-maximal interpretations for accounts that assume a maximal semantics for definite plurals.[2]

However, as Brisson (1998) argues in detail, non-maximal interpretations are not restricted to collective readings of verbal predicates, but also arise for distributive ones. In fact, the materials from the experiment to be presented here fall into this category, as they have color adjectives as their predicates. Intuitively speaking, it does not seem appealing to say that some of the circles in a display, say, were black on behalf of the entire plural individual consisting of all of the circles in the display. So this particular path to allowing non-maximal interpretations, based on collective interpretations of predicates, will not be of any help in interpreting the results.

There are, however, a number of further options for allowing non-maximal interpretations while assuming a maximal semantics, all of which involve some type of pragmatic mechanism that weakens the universality expressed in the semantics. First, this can be done by restricting the domain of the plural definite, along the lines of standard domain restriction in noun phrases (Westerstahl, 1984; von Fintel, 1994). Secondly, it can be done by restricting the domain of the distributive operator, e.g., in the form proposed by Brisson (1998), who uses the cover variable of Schwarzschild (1996) for this purpose (see also Brisson, 2003). Finally, a different pragmatic perspective has been suggested by Lasersohn (1999), who provides a formal implementation of the intuitive notion of 'pragmatic slack', which captures the fact that with a variety of constructions we are willing to allow a certain amount of imprecision based on the context they are used in. All three of these

[2]A wide range of proposals with different notions of what the relevant plural individual is have appealed to this account of non-maximal interpretations. See, for example, Landman (1996); Lasersohn (1995); Schwarzschild (1996).

approaches have in common that they assume that the semantics itself introduces strictly universal requirements for definite plurals combining with distributive predicates, but allow for some form of 'pragmatic weakening', as I will call it, which makes non-maximal interpretations possible. We will provide some more detailed aspects of these approaches in connection with the discussion of the experimental results below.

2.2. 'Weak' Accounts of Plural Definites

The 'strong' accounts we just reviewed see the basic semantic meaning of a plural definite as requiring maximality, which can be weakened on pragmatic grounds. Another family of accounts, call them the 'weak' accounts of plural definites, sees things exactly the other way around: they assume that the literal semantic contribution of a plural definite essentially is equivalent to existential quantification, whereas the maximality part only comes in as part of additional steps of pragmatic reasoning when needed.

While simple existential proposals along these lines have not, to my knowledge, been fully fleshed out in the literature, they have been hinted at by several authors in connection with a parallel analysis of singular definites. For example, Szabó (2000) proposes a semantics for singular definites as merely involving existential quantification. He supplements this with a pragmatic version of a file-card semantics Heim (1982), which then also helps to derive certain uniqueness effects. It would only seem natural to extend this line of thought to plural definites, and, in fact, Szabo notes in a footnote that he 'think[s] 'the Fs are G' is semantically closer to 'Some (contextually relevant) Fs are G" (Szabó, 2000, p. 52, footnote 2). A tentative extension of his account to plural definites is spelled out in Brogaard (2007, though note that she does not endorse this view). In a similar vein, Ludlow and Segal (2004), assume an existential semantics for definite descriptions, and derive uniqueness in terms of a process of Gricean reasoning. While it may not be entirely clear how to spell out an extension of their account to plural definites (again see Brogaard, 2007, for some discussion), the relation between what is semantically encoded and what is pragmatically derived for any such account would fall into the category of 'weak' accounts.

A more recent proposal that would seem to fall into this same category is that of Malamud (2012). Upon reviewing a variety of proposals for plural definites, Malamud finds all of them lacking in that they do not provide any details regarding the contextual factors that clearly affect maximality with plural definites. She presents a detailed decision-theoretic framework to capture such pragmatic factors, in particular the interlocutors' goals. As far as the semantics is concerned, Malamud builds on a cover-based account, in particular the variant of Schwarzschild's theory that Landman (1996) introduces as 'Theory IV'. While it does make use of covers, however, the account she adopts does not see these as being supplied by the context. Rather, it assumes that as far as the literal semantics is concerned, a statement with a plural definite simply gives rise to an existential requirement, namely that there be some cover under which the statement is true. As Malamud

herself notes, this requirement is extremely weak, since it essentially reduces to an existential analysis of plural definites. However, she sees this as a virtue, because her proposal takes what she calls an underspecified meaning with regards to maximality as the basic starting point, with decision-theoretic pragmatics doing the additional work in figuring out what type of interpretation is most appropriate in a given context.

What differentiates the weak and the strong theories, then, is whether or not maximality is part of the semantics of a plural definite. For strong theories, it is, and non-maximal interpretations are derived pragmatically. For weak theories, the semantics does not include a maximality requirement, but such a requirement can be added in at the level of pragmatic reasoning.

3. Experiment on Maximal vs. Non-maximal interpretations of Definite Plurals

The general approach taken in the present experiment builds on the by now fairly extensive literature on semantic and pragmatic processing. Most relevantly, a substantial body of work has argued, based on a variety of processing measures, that literal content, encoded at the level of lexical semantic entries for specific expressions, is available earlier in online processing than pragmatic aspects of meaning, such as scalar implicatures (Grice, 1975). While these results are not uncontested, it is nonetheless plausible in general to assume that literal meanings are more or less immediately available in processing, and that pragmatic considerations that draw directly on those literal meanings, as well as a variety of contextual information, are secondary both in theoretical terms and in terms of the temporal sequence of cognitive steps in actual processing.

Applying this line of thinking to the phenomenon at hand, and in particular the relation between maximal and non-maximal interpretations of definite plurals, the two types of accounts sketched above differ in the predictions they make for online processing. On the strong account, the literal meaning is a maximal one, whereas non-maximal interpretations are derived pragmatically. We would thus expect the former to be faster than the latter from this perspective. The reverse holds from the perspective of the weak accounts: these assume that the literal meaning of statements with plural definites either involve (the equivalent of) existential quantification or are underspecified with respect to maximality.[3] Therefore, non-maximal interpretations are based on the literal meaning alone, whereas maximal ones require some form of additional pragmatic reasoning. This predicts, at least, that maximal interpretations should not be faster than non-maximal ones, and - depending on further assumptions - might even predict maximal ones to be slower.

[3]It is actually not entirely clear to me that the latter option can be differentiated from the former, since the literal meaning that Malamud considers still is an existential one. But we'll give the account the benefit of the doubt for the moment, since the predictions will diverge from the those of the strong account either way.

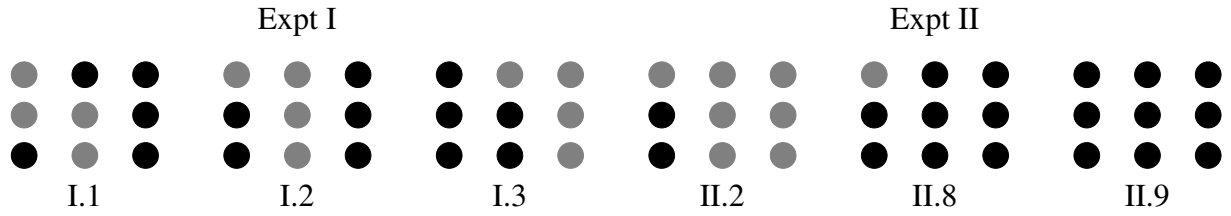

Figure 1: Experimental Displays

3.1. Methods

Design In order to keep things simple, and furthermore to avoid the potential confound of non-maximal interpretations being due to a collective interpretation, the experimental task involved truth value judgments relative to simple arrays of colored shapes. Each display contained nine shapes of the same type. The shapes in each display generally (except for one control condition) came in two colors. The distribution of the colors was varied, both in terms of the number of shapes with a given color and with regard to the relative placement of the different colored shapes. The goal in introducing these variations was to assess, in addition to the reaction times for maximal and non-maximal interpretations, the types of pragmatic factors that affect the likelihood of non-maximal interpretations in the first place. Furthermore, the sentences to evaluate were varied in terms of whether they applied to the entire display or to some part of it. The latter was implemented by adding a locational PP modifier to the noun phrase for the shape. For purposes of illustration of the various conditions, it is convenient to group them into four sub-experiments, but note that comparisons between conditions across sub-experiments are just as valid as comparisons within each sub-experiment, since all the data was collected together and the nature of the stimuli resulted in a very minimal variation of the materials.

One consideration in designing the specifics of the task was to make the experience a bit more challenging and interesting, and to draw subjects' attention away from the fact that they may sometimes be put in a position where they have to judge sentences that are hard to evaluate once you start thinking too much about them (they also saw sentences with singular definites that did not successfully pick out any item in the relevant display). To achieve this, we incorporated a simple memory component into the task. Subjects were shown the displays of colored shapes for a brief period of time (600ms), and then read the sentence they had to evaluate. The length of presentation was chosen so as to not make the task too difficult, but still taxing enough that subjects were unlikely to engage in any prolonged deliberations about what the sentences they were seeing could mean upon reflection.

The various types of displays are provided in Figure 1. The sentences shown were as follows:

(6) **NoPP** The circles were black.

PP The circles on the left were black.

Let us begin by considering the first set of display variations in light of the **NoPP** sentence. The number (and proportion) of black circles was kept constant across all displays, namely five (out of nine). What was varied was the extent to which the black circles appeared in a contiguous spatial configuration: In I.3, all of them were contiguous at the left edge of the display, In I.1, four out of the five black circles appeared next to each other. And in I.2, there was a split of two and three black circles appearing together in different areas of the display. The manipulation here was partly based on the raw intuition that the extent to which the circles displaying the relevant property (here, having the same color) form a salient sub-part of the display has a clear impact on ones likelihood of accepting the sentence 'the circles were black' as true. From a theoretical perspective, this can also be related to the domain restriction approach to non-maximal interpretations (see below).

Turning to the **PP**-version of the sentence, the variation in the displays now resulted in a variation of the number (and proportion) of individuals in the NP denotation that had the property predicated of them. In particular, I.1 only had one circle on the left that was black, I.2 had two, while I.3 had all 3 circles be black. Manipulating the number and proportion of the shapes that would fall under the predicate clearly also seems relevant to the likelihood of providing a non-maximal judgment in intuitive terms. In fact, claims along these lines have been made in the literature, e.g., by Brisson (1998, p. 49), and Brogaard (2007), who claims that 'the smaller the number of individuals in the domain the more likely it is that all the individuals are taken to satisfy the predicate (individually or collectively)' (p. 419).

The second set of displays are provided on the right side of Figure 1. The sentences were the same as above, in the **NoPP** and **PP** versions. Beginning with the former, which again applied to the entire display, we now are looking at a further variation of the number and proportion of circles that have the relevant color (here: black). Display II.2 contained two such circles, and seven circles of another color (here: gray). Display II.8 had eight black circles and one gray one. And Display II.9 served as a control with all nine circles in black. Together with the other displays, this gives us a fairly large spread of different proportions of shapes satisfying the color predicate, from 22-89% in the non-maximal conditions.

The **PP** condition didn't add further variations to the number or proportion dimension, as the circles on the left matched the ones from display I. However, what was varied here was the color of the other shapes, again largely based on an intuition, namely that the color of the other shapes might affect the availability of the non-maximal interpretation: if none of the other shapes are black, then relative to that, two out of three shapes being black might suffice to give the non-maximal judgment that it's true that the circles on the left were black. But if the others are all black, than relative to that the circles on the left may not seem to display a sufficient degree of blackness.

To sum up the main features of the design, we set out to test reaction times for a variety of conditions for both maximal and non-maximal interpretations of plural definites. Further variations explored various possible pragmatic factors that might have an impact on the availability of non-maximal interpretations. In particular, we are testing whether the clustering of the shapes of the relevant color plays a role, whether the number and proportion affects non-maximal interpretations, and whether something like a standard set by other shapes present in the display might affect the willingness to give a non-maximal judgment.

Materials Using a set of six shapes (circles, crosses, diamonds, hearts, squares, and triangles) and six colors (black, blue, gray, green, red, and yellow), 72 items with versions in all twelve conditions described above were created and split into six lists containing six items per condition. This created a fully counter-balanced design, where each subject saw six items in each condition, and each item was seen in all conditions by different subjects. The location of the critical shapes was varied systematically and the location specified in the presuppositional phrases was adjusted accordingly. In addition to the 72 experimental items, each list contained 90 other items from other experiments. 72 of these involved singular definite statements of the form 'The circle on the left was black' with slightly different displays (containing only five colored shapes), which varied systematically in whether the prepositional phrase and the color adjective were true of a circle in the display (see Schwarz, 2012, for discussion of this experiment). Another 18 items contained arrays similar to the ones presented here, followed by sentences containing the quantifiers 'few' or 'a few'. The order of presentation of items within each list was randomized for each subject, with no more than two subsequent trials from the same sub-experiment.

Procedure & Participants The experimental design was implemented using the Experiment Builder software package by SR Research and responses and reading- and reaction-time data were recorded using an EyeLink 1000 eye tracker. Subjects were seated in front of a computer screen and received the initial instructions below, followed by a practice trial.

(7) **Instructions**
In this experiment, you will be shown simple pictures containing various shapes for a brief period of time. Afterwards, you will see a sentence, and your task is to evaluate whether the sentence is true or false relative to the displayed array. Try to push the appropriate button as quickly as possible.
We will begin with a brief practice trial. Then we have to set up the eye tracker.
If you have any questions, please feel free to ask the experimenter now!
When you are ready, press a button to proceed to the practice trial.

Each trial consisted of the following sequence of events:

(8) Sequence of events during each trial
   a. Display of dot in center to control for initial eye position
   b. Display of array of colored shapes for 600ms
   c. 10ms pause
   d. Display of dot in center to control for initial eye position
   e. Display of sentence
   f. Button press to indicate 'true'/'false' answer

Responses were recorded using a ResponsePixx button box with 5 buttons. The left and right buttons were labeled as 'true' and 'false', with position of these values counter-balanced between subjects. After participating, subjects received a debriefing with a short explanation of the purpose of the experiment. 48 undergraduate students at the University of Pennsylvania participated for class credit.

3.2. Results

Response Data Responses from the TVJ-task were coded as maximal vs. non-maximal. For all conditions where only some of the shapes in question had the color expressed by the adjective, a 'true' response was coded as non-maximal, and a 'false' response as maximal. For the conditions where the relevant shapes had the same color, 'true' responses corresponded to maximal interpretations, though they are strictly speaking also consistent with a non-maximal interpretation. 'False' responses in these cases are errors (since neither interpretation would support such a response), and thus give us an indication of the general error rate incurred by the task setup.

Overall, there was a substantial number of non-maximal responses. The mean for all data from conditions with a non-maximal number of circles (greater than 1)[4] of the right color (all displays from Expt I in the **NoPP**-condition, the I.2-**PP** condition, and both the **PP** and **NoPP** conditions for displays II.2 and II.8) was 70% maximal response choices, and 30% non-maximal response choices. While there clearly were more maximal responses than non-maximal ones, the non-maximal ones are unlikely to be due to error alone, given the high accuracy in the maximal conditions (>97% maximal choices in conditions II.9-NoPP and II.9-PP, and $>$ 90% in the I.3-PP condition, where error was presumably higher due to having to remember the PP, as well as the corresponding details of the display, correctly). Furthermore, accuracy was similarly high in another sub-experiment, using sentences such as *(A) Few of the circles were black* with displays similar to II.2, which also indicates that remembering the distribution of colors correctly was not so hard as to give rise to a 30% error rate.

[4]Condition I.1-PP, which had a single black circle on the left seemed to behave slightly differently, presumably due to the additional issue of the noun phrase being plural. The frequency of non-maximal responses for it were on par with the maximal control condition I.3-PP, at about 10%. We do not include this condition in further analyses.

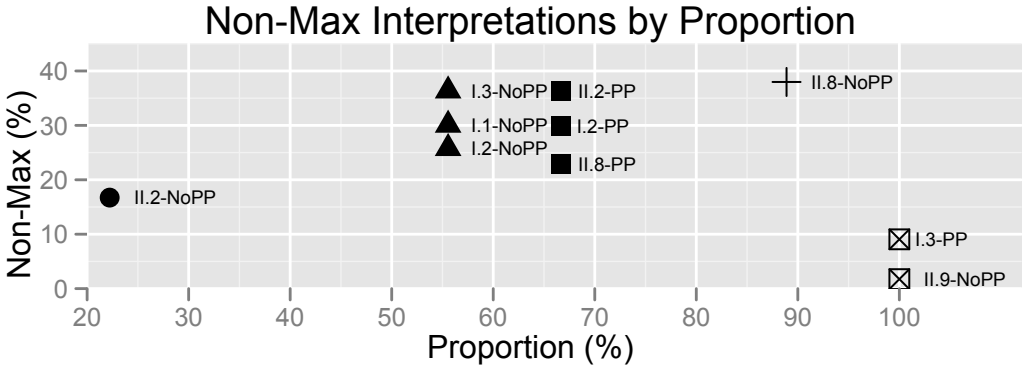

Figure 2: Non-maximal interpretations (error for maximal control conditions) by proportion of relevant shapes of the right color

A direct statistical comparison between the frequency of non-maximal interpretations in the maximal and non-maximal conditions (as characterized above), implemented by carrying out a logistic regression using the lmer function from the lme4-package in R Bates (2005), with random intercepts for subjects and items, as well as slopes for the factor NonMax,[5] yielded a significant difference (means: 29.5% for NonMax and 4.3% for Max; $p's < .001$).

Turning to a more detailed perspective on the specific manipulations between conditions, we carried out a number of planned comparisons, using logistic regressions as above on subsets of the data. The means for all relevant conditions[6] are plotted in Figure 2, ordered on the x-axis by proportion, which we turn to in more detail below.

To begin with, the Expt I-NoPP conditions looked at possible effects of clustering or contiguity. We found differences in frequency of non-maximal interpretations, with the greatest value in the fully contiguous condition I.3-NoPP (36%), an intermediate value in condition I.1-NoPP (30%), and the smallest in condition I.1-NoPP (26%). Using the fully contiguous condition I.3-NoPP as a reference level in the analysis, the differences between it and the I.1-NoPP condition as well as the I.2-NoPP condition were significant ($p < .05$ and $p < .001$ respectively). The difference between I.1-NoPP and I.2-NoPP was not significant when including random slopes in the model, and only marginally significant ($p < .1$) when only random intercepts were included. In sum, a fully

[5]Following recent arguments by Barr et al. (in press) that maximal random effect structures should be used when possible, we generally computed models with the maximal random effect structure that would converge, with random effect slopes for each factor (and an interaction where applicable). However, we also carried out parallel analyses with only random intercepts to insure against the risk of over-fitting. In the following, unless otherwise noted, both types of analyses yielded the same result.

[6]II.9-PP is omitted for ease of presentation; numerically, it was almost exactly identical to II.-NoPP; I.1-PP is left out for the reasons discussed above.

contiguous distribution of black circles yielded a significant increase in non-maximal responses.

The second type of manipulation concerned the color of the other shapes: conditions II.2-PP and II.8-PP both had two black circles on the left, but the former had all other circles in gray, and the latter in black. This too had an effect on the proportion of non-maximal interpretations, with means of 36% in condition II.2-PP and 23% in condition II.8-PP. The difference between these two was revealed to be significant by a logistic regression analysis, and they both were also significantly different from condition II.9-PP (for both the fullest converging model with a random slope for subjects, and for a model with random intercepts for subjects and items only; for the latter, all $p's < .001$).

Turning to the analysis of proportion as a factor, a visual inspection of the graph in Figure 2 already suggests that proportion does not have a large effect on the availability of non-maximal interpretations, in particular in comparison with the effects of the other factors we have considered. The conditions where we had maximal clustering (I.3-NoPP) and where the other shapes were not black (II.2-PP), with black-circle proportions of 55% and 66% respectively, did not display a significantly lower proportion of non-maximal interpretations than condition II.8-NoPP with a black-circle proportion of 89%. However, all three of these conditions differed significantly from II.2-NoPP, with a black-circle proportion of 22% ($p < .05$ for I.3-NoPP comparison, other $p's < 0.001$). This latter condition still did differ from the control condition, in that it displayed more non-maximal response choices ($p < .01$). Regressing response choices in all non-maximal conditions on proportion alone did yield a significant effect, but the size of the effect estimate was quite small. Excluding condition II.2-NoPP, which may have been a border-line case in terms of proportion, and including factors for clustering and the color of other shapes, resulted in the effect of proportion becoming non-significant. In sum, while proportion may play a role in the availability of non-maximal interpretations at the low end of the proportion-range, there is no, or at best a very small, effect of proportion in the range from 50-90% when other factors are accounted for in the model.

Reaction Times Reaction times were analyzed as the time that passed between the initial display of the sentence to be judged and the button press indicating the True/False-response. An overview for all conditions, split by the type of response (with Non-max indicating erroneous responses in the maximal conditions) is provided in Figure 3. As can be seen very clearly in the **NoPP** conditions, non-maximal responses took longer than maximal ones. To establish the statistical significance of this difference, a number of mixed-effect model analyses with subjects and items as random effects, using the $lmer$ function of the lme4 package in R (Bates, 2005), were carried out. For models with a simple random effect structure that only included random intercepts, p-values were calculated using MCMC estimates for significance (Baayen et al., 2008). We report t-values for models with the maximal converging random effect structure as well.[7]

[7]MCMC simulations are not yet implemented for complex random effect structures in lme4; given the size of our data set, t-values greater than 2 roughly correspond to significance at the conventional $\alpha = .05$ level.

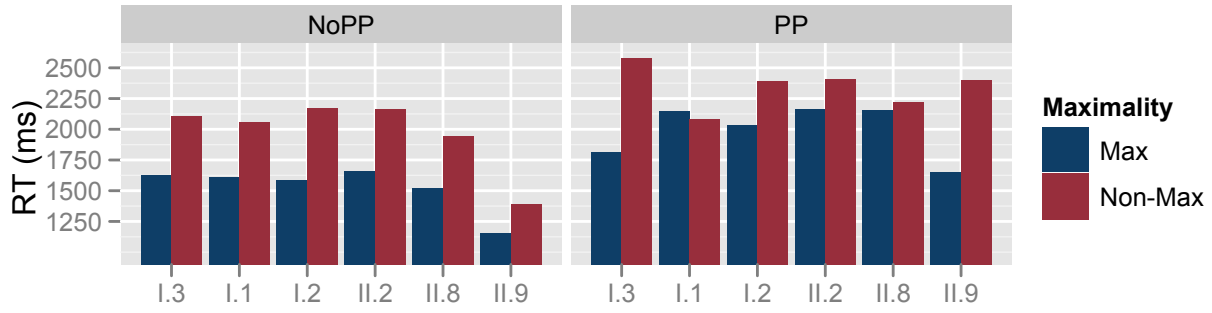

Figure 3: Reaction times by response and condition (Non-Max represents response errors for maximal conditions (II.9-NoPP, I.3-PP, II.9-PP)

Regressing reaction times on the type of response (maximal vs. non-maximal) for all non-maximal conditions together (I.1-NoPP, I.2-NoPP, I.3-NoPP, I.2-PP, II.2-NoPP, II.8-NoPP, II.2-PP, II.8-PP) yielded a significant difference ($t = 5.72$, $p < .001$ for simple random effect model with random intercepts for subjects and items only; $t = 3.50$ for maximal model with random slopes for subjects and items), with means of 2175ms for non-maximal responses and 1795ms for maximal ones. Note that non-maximal responses corresponded to a 'true'-judgment, and maximal ones to a 'false'-judgement, so the difference here likely is even larger if there is any response-bias for 'true'-responses being faster, as frequently reported in the literature.

Analyzing differences in reaction times for subsets of the data also yielded significant results for a variety of conditions. The Expt I displays in the **NoPP** condition all had significantly faster reaction times for non-maximal responses than for maximal ones, as did condition I.2-PP (all $p's < .05$). Conditions II.2-NoPP and II.8-NoPP together also displayed the same significant difference, and so did condition II.2-PP ($p's < .05$).

In sum, there is ample evidence from the reaction time analysis that subjects took more time when making non-maximal response choices than when making maximal response choices.

Summary of Results The analysis of our experimental data shows that there is a substantial proportion of non-maximal response choices, though maximal ones are clearly preferred overall. Both of the two main manipulations, concerning the clustering of the shapes of the right color and the presence or absence of contrast with the color of the other shapes in the display had a significant effect on the frequency of non-maximal response choices, suggesting that there are pragmatic factors at play in the process of settling on a decision. The analysis of the reaction time data furthermore provided evidence that maximal responses were generally faster than non-

maximal ones, despite the fact that the former corresponded to 'false' judgments, which are often found to be slower than 'true' judgments on independent grounds in the literature. The additional time incurred prior to providing a non-maximal response suggests that the process of reaching the response decision involves additional pragmatic considerations that add to the time that would be required to give a response on semantic grounds alone. Both the response distribution and the reaction time data thus lend support to 'strong' accounts of plural definites, which see them as encoding maximality on the semantic level, with certain pragmatic avenues open for reaching a non-maximal interpretation in certain circumstances.

4. General Discussion

The overall results of increased reaction times for non-maximal interpretations together with pragmatic factors modulating the availability of the non-maximal interpretations lend clear support to accounts that assume a maximality-based semantics. Unlike in previous studies (e.g. Caponigro et al., 2012), which mostly found chance level results in TVJ-tasks, our TVJ-task yielded a majority of maximal interpretations.[8] The differences in reaction times further speak against understanding the response decision as a random choice. The difference in response frequencies between maximal and non-maximal conditions speak against seeing non-maximal interpretations as errors throughout, as does the variation based on the various distributional manipulations across conditions. The response results also do not seem consistent with the view that statements with definite plurals and distributive predicates generally involve a homogeneity presupposition (Löbner, 2000), as there seems to be a limited, but genuine availability of non-maximal interpretations, which -again - seems to be modulated further by a variety of pragmatic factors.

Given the delay for non-maximal responses and the modulation of the availability of non-maximal interpretations by pragmatic factors, the results do not seem consistent with the picture provided by 'weak' accounts of plural definites as discussed above. These would assume that non-maximal interpretations are what is semantically encoded, and thus would lead us to expect that corresponding responses should be faster than ones that require further reflection and consideration of pragmatic factors. Furthermore, even though Malamud (2012) provides an intriguing and detailed analysis of how pragmatic factors, particularly the goals of the speakers, can influence the extent to which maximality enters the picture, it seems unlikely that the variation across conditions found here could be explained in terms of her analysis, since the task and task demands remained constant across all trials and conditions.

If we take the data reported here to support a 'strong' account, which assumes a maximal semantics combined with the possibility of pragmatic weakening, we should at least begin to drill a bit deeper and consider the options for spelling out such a perspective in some more detail. The general idea throughout all possible versions would be that the initial, and most available interpretation of the stimuli in our experiment involves the maximal literal semantics paired with the most obvi-

[8] Statistically, this can be supported, e.g., by a base-line model with only random effects, which yields a significantly positive intercept ($p < .001$).

ous pragmatic choice, e.g., taking the entire display as the relevant frame of reference. Additional pragmatic considerations may enter the picture to adjust the extent of maximality, perhaps to support a desire for charitable interpretations that allow 'true' responses. As mentioned above, we can consider various implementations of such an account, either by invoking some type of domain restriction, or an analysis of 'pragmatic slack' more generally. The first can be implemented in various ways.

First, it is commonly assumed that quantificational noun phrases involve implicit restriction of the domain that they are applying to, and that this has to be done for each noun phrase separately (Westerstahl, 1984; von Fintel, 1994). Based on such accounts, we could then assume that non-maximal answers are really based on a maximal semantics, but that maximality only has to hold within a restricted domain that does not encompass all of the shapes in the display. However, more needs to be said on such an account so as to account for the differences between definite plurals and other quantifiers (since no comparable weakening is possible with universal quantifiers!), as well as to spell out more concrete predictions that will relate it to the specific manipulations considered here. One promising variant of this approach in light of the effect of clustering observed above would be a situation-based view of domain restriction (Cooper, 1995; Recanati, 1996; Kratzer, 2004; Schwarz, 2011). Clustering could facilitate the choice of an alternative domain on this view, because a contiguous group of like-colored shapes suggests itself as a salient alternative situational sub-domain, relative to which maximality holds.

Another variant of a domain restriction approach considers the relevant restriction to be introduced at the level of the distributive operator. In particular, building on Schwarzschild (1996), Brisson (1998) makes use of the notion of covers and assumes that they provide the domain restriction for the universal quantification introduced by the distributive operator. A cover of a set is a set of subsets where each element of the original set is an element of at least one such subset. Unlike with the related notion of a partition, the subsets are allowed to overlap. Using covers to modulate the quantificational effect of the distributive operator makes it possible to capture cases of intermediate distributivity (Schwarzschild, 1996). However, most crucially for our purposes, covers can give rise to domain restriction effects with definite plurals, which in turn account for non-maximal interpretations. To illustrate, consider the following example from Brisson (1998, p. 82):

(9) a. *The boys are hungry.*
b. $\forall x[x \in [\![Cov_i]\!] \,\&\, x \subseteq [\![\textit{the boys}]\!] \rightarrow x \in [\![\textit{hungry}]\!]$

While there still is universal quantification introduced by the distributive operator, there are additional constraints on this quantification due to the cover. In particular, quantification only applies to those subsets of the denotation of *the boys* that form an element of the cover. One use of this is, as noted above, to account for intermediate distributivity. But it also allows us to have certain boys drop out of the picture, namely if they appear only in elements of the cover that also contain non-boys. Brisson calls such covers, where no set of cells of the cover has as its union the set

of boys, 'ill-fitting'. While Schwarzschild had proposed to rule out such covers as pathological, Brisson argues that they are just what we need to account for non-maximal interpretations of plural definites with distributive predicates. Thus, the notion of a cover provides another way of introducing domain restriction for plural definites, and it seems equally appealing to account for the clustering effect by assuming that the salient contiguous grouping of like-colored shapes helps to suggest just the right ill-fitting cover.[9]

Finally, there is a rather different perspective on non-maximal readings of plural definites, due to Lasersohn (1999), which relates it to a more general notion of 'pragmatic slack'. The initial observation is that there are all kinds of phenomena where, depending on context and the purposes of conversation, sentences that are strictly speaking false can be regarded as true because at the level of detail that is deemed relevant, they should just as well count as true. For example, in most contexts it will be regarded as true that John arrived at 3pm, even if he actually arrived two seconds after 3pm. Similarly, a plural definite, seen as having a maximal interpretation, may be regarded true even if there are some exceptions, assuming those exceptions don't matter for present purposes. Various expressions, such as *exactly*, can be seen as 'slack regulators' from this perspective, and this also helps to differentiate between a simple definite plural statement (*the boys left*) and the corresponding statement with *all* (*all the boys left*). The meaning of *all*, on this account, precisely involves slack regulation, disallowing the pragmatic flexibilities displayed by plural definites. The latter, however, allow precisely for the 'pragmatic halos', as Lasersohn calls them, that can provide the type of pragmatic weakening involved in non-maximal interpretations of plural definites. Just how the availability of such halos is modulated by the experimental manipulations considered here is a question that will have to be left for further research.

In conclusion, let us step back and consider the broader picture concerning the semantics and pragmatics of plural definites. There is no doubt that there are more factors involved than we have considered here. For one, the existence of other non-maximal interpretations due to collectivity is of course by no means inconsistent with our findings. Furthermore, the arguments presented by Malamud (2012) for the role of interlocutors' goals seem rather strong, even though they don't seem to help us in accounting for the variation found in our data. However, it may well be worth trying to incorporate aspects of speaker's goals into an analysis of domain restriction, e.g., by involving the notion of a Question under Discussion (Roberts, 1996), as has been proposed by Schwarz (2009), building on Kratzer (2007), which might allow us to capture the various results in a uniform manner. Again, further details have to be left for future work. To end on a speculative note related to this, it is quite interesting that Caponigro et al. (2012) find that the developmental path towards maximal interpretations in acquisition seems rather slow. While one could see that as evidence against a semantically encoded maximality requirement, an alternative perspective, directly tied into the line of thinking based on nominal domain restriction considered above, would

[9]One potential problem for this approach is that it would require that the domain of entities relative to which the cover is considered has to include shapes other than the ones shown in a given display, because ill-fitting covers only can model domain restriction effects if there are entities in the domain that don't meet the description expressed by the NP in the plural definite.

be to hypothesize that what children are struggling with is the appropriate - or at any rate adult-like - choice of a domain of interpretation in the relevant experimental tasks. This opens up a host of interesting questions relating the developmental patterns of quantifier interpretation more generally.

References

Baayen, R. H., D. J. Davidson, and D. M. Bates (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language 59*, 390–412.

Bates, D. M. (2005). Fitting linear mixed models in r. *R News 5*, 27–30.

Brisson, C. (1998). *Distributivity, Maximality, and Floating Quantifiers*. Ph. D. thesis, Rutgers, New Brunswick, NJ.

Brisson, C. (2003, April). Plurals, 'all', and the nonuniformity of collective predication. *Linguistics and Philosophy 26*(2), 129–184.

Brogaard, B. (2007, September). The but not all: A partitive account of plural definite descriptions. *Mind & Language 22*(4), 402–426.

Caponigro, I., L. Pearl, N. Brooks, and D. Barner (2012). Acquiring the meaning of free relative clauses and plural definite descriptions. *Journal of Semantics 29*(2), 262–293.

Cooper, R. (1995). The role of situations in generalized quantifiers. In S. Lappin (Ed.), *Handbook of Contemporary Semantic Theory*. Blackwell.

von Fintel, K. (1994). *Restrictions on Quantifier Domains*. Ph. D. thesis, UMass.

Grice, H. (1975). Logic and conversation. In P. Cole and J. Morgan (Eds.), *Syntax and semantics: Speech acts*. New York: Academic Press.

Heim, I. (1982). *The semantics of definite and indefinite noun phrases*. Ph. D. thesis, UMass.

Kratzer, A. (2004). Covert quantifier restrictions in natural languages. talk given at Palazzo Feltrinelli in Gargnano June 11, 2004.

Kratzer, A. (2007). Situations in natural language semantics. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. Stanford: CSLI.

Landman, F. (1989, October). Groups, i. *Linguistics and Philosophy 12*(5), 559–605.

Landman, F. (1996). Plurality. In S. Lappin (Ed.), *THe Handbook of Contemporary Semantic Theory*, pp. 425–457. Blackwell, Cambridge.

Lasersohn, P. (1995). *Plurality, conjunction, and events*. Dordrecht; Boston: Kluwer Academic Publishers.

Lasersohn, P. (1999). Pragmatic halos. *Language 75*(3), 522–551.

Link, G. (1983). The logical analysis of plurals and mass terms: A lattice-theoretical approach. In R. Bäuerle, C. Schwarze, and A. von Stechow (Eds.), *Meaning, Use, and Interpretation of Language*, pp. 303–23. Berlin: de Gruyter.

Löbner, S. (2000, April). Polarity in natural language: Predication, quantification and negation in particular and characterizing sentences. *Linguistics and Philosophy 23*(3), 213–308.

Ludlow, P. and G. Segal (2004). On a unitary semantical analysis for definite and indefinite descriptions. In A. Bezuidenhout and M. Reimer (Eds.), *Descr*. Oxford: Oxford University Press.

Malamud, S. A. (2012, July). The meaning of plural definites: A decision-theoretic approach. *Semantics and Pragmatics 5*(0), 3:1–58.

Neale, S. (1990, March). Descriptive pronouns and donkey anaphora. *The Journal of Philosophy 87*(3), 113–150.

Recanati, F. (1996). Domains of discourse. *Linguistics and Philosophy 19*(5), 445–475.

Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theor of pragmatics. In J. H. Yoon and A. Kathol (Eds.), *Papers in Semantics*, Number 49 in OSU Working Papers in Linguistics, pp. 91–136. OSU.

Schwarz, F. (2009). *Two Types of Definites in Natural Language*. Ph. D. thesis, University of Massachusetts Amherst, Amherst, MA.

Schwarz, F. (2011). Situation pronouns and domain restriction. Ms.

Schwarz, F. (2012). Rejecting statements with non-referring definites – experimental support for a presuppositional view. Ms., UPenn.

Schwarzschild, R. (1996). *Pluralities*. Dordrecht; Boston: Kluwer Academic.

Sharvy, R. (1980, October). A more general theory of definite descriptions. *The Philosophical Review 89*(4), 607–624.

Strawson, P. (1950). On referring. *Mind 59*, 320–344.

Szabó, Z. G. (2000). Descriptions and uniqueness. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition 101*(1), 29–57.

Westerstahl, D. (1984). Determiners and context sets. In J. van Benthem and A. ter Meulen (Eds.), *Generalized Quantifiers in Natural Language*, pp. 45–71. Foris.