Nat Lang Semantics (2015) 23:119–156
DOI 10.1007/s11050-015-9113-0

Experimental investigations of ambiguity: the case of *most*

Hadas Kotek · Yasutada Sudo · Martin Hackl

Published online: 25 April 2015

Abstract In the study of natural language quantification, much recent attention has been devoted to the investigation of verification procedures associated with the proportional quantifier *most*. The aim of these studies is to go beyond the traditional characterization of the semantics of *most*, which is confined to explicating its truth-functional and presuppositional content as well as its combinatorial properties, as these aspects underdetermine the correct analysis of *most.* The present paper contributes to this effort by presenting new experimental evidence in support of a decompositional analysis of *most* according to which it is a superlative construction built from a gradable predicate *many* or *much* and the superlative operator *-est* (Hackl, in Nat Lang Semant 17:63–98, 2009). Our evidence comes in the form of verification profiles for sentences like *Most of the dots are blue* which, we argue, reflect the existence of a superlative reading of *most*. This notably contrasts with Lidz et al.'s (Nat Lang Semant 19:227–256, 2011) results. To reconcile the two sets of data, we argue, it is necessary to take important differences in task demands into account, which impose limits on the conclusions that can be drawn from these studies.

Keywords Quantification · Superlatives · Experimental design · Language processing · Semantics–cognition interface · *Most*

H. Kotek (✉)
Department of Linguistics, McGill University, 1085 Dr. Penfield, Montreal, QC H3A 1A7, Canada
e-mail: hkotek@alum.mit.edu

Y. Sudo
Division of Psychology and Language Sciences, Faculty of Brain Sciences, University College London, Chandler House, Room 115H, 2 Wakefield Street, London WC1N 1PF, UK

M. Hackl
Department of Linguistics and Philosophy, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA

1 Introduction: *most* in subject position

It is well known that *most* in object position has both proportional uses, in which case it appears in its bare form and can be reasonably well paraphrased by *more than half*, as in (1a), and superlative uses similar to the reading that is obtained when *most* combines with certain kinds of degree predicates (e.g., *most expensive car*). In the latter cases, *most* is accompanied by the definite article, as in (1b) (Bresnan 1973).[1]

(1) a. John talked to *most of the* students. *proportional*
≈ John talked to more than half of the students
b. John talked to the *most* students. *superlative*
≈ John talked to more students than anybody else

There have been several attempts in the literature to relate the two uses of *most* (e.g., Pinkham 1985; Yabushita 1999; Hackl 2009; Krasikova 2011; Szabolsci 2012). The canonical view among them is that this is a case of lexical ambiguity, with bare *most* a quantificational determiner (Barwise and Cooper 1981) and *the most* a superlative construction (e.g. Szabolcsi 1986).[2]

When *most* occurs in subject position, only bare *most* is grammatical. *The most* is degraded and even ungrammatical for many speakers, as illustrated by (2).[3]

(2) a. *Most* of the students talked to John.
b. ??The *most* students talked to John.

The range of possible interpretations available to bare *most* in subject position as well as the associated verification strategies have become a central topic of debate in recent years (e.g. Hackl 2009; Pietroski et al. 2009, 2011; Lidz et al. 2011; Solt 2011; Kotek et al. 2011a, 2011b). More specifically, while it is uncontroversial that bare *most* in subject position can give rise to proportional truth conditions similar to the ones expressed by *more than half,* it has been argued in Hackl (2009) that the verification strategy associated with it is quite different from the one associated with

[1] Bare *most* can also occur outside of the partitive frame, as in (i) below, and in that case the sentence tends to have a generic interpretation. See Matthewson (2001) and Szabolsci (2012) for a discussion of such cases.

(i) Most linguists are millionaires. (Matthewson 2001, ex. 47)

[2] See Szabolcsi (2010) for a review of the history of this debate.

[3] The extent to which *the most* in subject position is degraded is subject to grammatical factors as well as dialectal variation. For instance, data like (i) from Kotek et al. (2011b), are acceptable to most native speakers of English.

(i) Where do the most students live?

more than half and reflects its superlative morphosyntax. Moreover, Kotek et al. (2011a) have argued that bare *most* in subject position is in fact ambiguous between a (preferred) proportional reading and a (latent) superlative reading. This contrasts markedly with the theoretical framework of Lidz et al. (2011), who conduct a sentence verification study of bare *most* to adjudicate between different ways of describing the truth-conditional import of this form. Lidz et al. assume that bare *most* is unambiguously proportional and, indeed, their results provide no indication that bare *most* might share any properties with *the most*—which, we should note, they do not explicitly discuss. Consequently, Lidz et al. propose an analysis of bare *most* that makes the prediction that bare *most* in subject position, like elsewhere, only gives rise to proportional truth conditions.

The present paper, by contrast, presents novel experimental evidence which supports an analysis of *most* according to which *the most* and bare *most* are built from the same basic ingredients—a gradable predicate *many/much* and the superlative operator *-est*—but project different LFs, as argued for e.g. in Hackl (2009). It also argues that the seemingly conflicting set of empirical results from previous studies can be reconciled with this structural view of *most* once the differences in experimental techniques employed by the various researchers are properly factored in. The paper is composed as follows: In Sect. 2 we briefly review the debate between an approach that views *most* as a lexical primitive and one which views *most* as a complex superlative construction. In Sect. 3 we present novel data from two experiments that support the view that *most* is uniformly a superlative construction. Section 4 discusses the implications of our results for the theory of *most*, differences between the determiners *most* and *more than half* that arise from the experiments, and also previous work on *most*—notably the work of Lidz et al. (2011), who conducted a very similar experiment but reported very different findings about the nature of *most* than we do here. We address several concerns regarding the experiment in Lidz et al. (2011) and the conclusions that were drawn based on this work.

2 Background: two theories of *most*

As we saw in (1a,b), *most* in object position has *proportional* and *superlative* uses. The availability of these two readings seems to correlate with the presence or absence of the definite article: bare *most* gives rise to a proportional reading, while *the most* produces a superlative reading.

The canonical take on these facts, implicit e.g. in Barwise and Cooper (1981), is to assume that the observed pattern is akin to a lexical ambiguity. That is, there are two unrelated *most*s: bare *most* is a lexical determiner while *the most* is a superlative construction built from a gradable predicate such as *many* or *much* and the superlative morpheme *-est*. A recent version of this view on bare *most*, proposed in Lidz et al. (2011), is expressed in (3). On this analysis, bare *most* is analyzed as a quantificational determiner that takes two set-denoting expressions A and B and yields True only if the number of As that are Bs is greater than the number of As minus the number of As that are Bs. This holds just in case there are more As that are Bs than there are As that are not Bs or, equivalently, more than half of the As are Bs.[4]

(3) a. John talked to most of the students.
b. $[\![\text{most}]\!](A)(B) = 1$ iff $|A \cap B| > |A| - |A \cap B|$
c. $[\![\text{John talked to most of the students}]\!] = 1$ iff $|\{x\text{: x is a student}\} \cap \{x\text{: John talked to x}\}| > |\{x\text{: x is a student}\}| - |\{x\text{: x is a student}\} \cap \{x\text{: John talked to x}\}|$

The superlative reading of *the most,* by contrast, is canonically analyzed as a construction that involves degree quantification, with the superlative morpheme -*est* denoting a degree quantifier that is restricted by a comparison class C, which, in the case of (4a), contains contextually salient alternatives to John. (4a) is true only if there is a plurality of students that John talked to that is more numerous than any plurality of students talked to by any contextually salient individual different from John (Heim 1985; Szabolcsi 1986).[5]

(4) a. John talked to the most students.
b. $[\![\text{John talked to the most students}]\!] = 1$ iff $\exists d \exists X[\text{students}(X)$ & John talked to X & $|X| \geq d$ & $\forall y \in C[y \neq \text{John} \rightarrow \neg\exists Y[\text{students}(Y)$ & y talked to Y & $|Y| \geq d]]]$

On the structural view, both the proportional and the superlative readings of *most* are analyzed as superlative constructions. The proposal in Hackl (2009) analyzes the superlative reading of *most* essentially as in (4) above. The proportional reading, on the other hand, is analyzed as in (5).[6]

(5) a. John talked to most of the students.
b. $[\![\text{John talked to most of the students}]\!] = 1$ iff $\exists d \exists X[\text{students}(X)$ & John talked to X & $|X| \geq d$ & $\forall Y \in C\ [\text{students}(Y)$ & $Y \perp X] \rightarrow |Y| < d]]$

(5b) is parallel to (4b) except that (i) the comparison class C is assumed to be the set of student pluralities rather than the set of contextually relevant people who talked to students and (ii) non-identity is assumed to hold between any two alternatives in C if they are non-overlapping pluralities of students. The symbol $\perp$ is used to

[4] Traditionally, the semantics of *most* is described as in (i) below, which is truth-conditionally equivalent to (3b). We discuss this notation and the difference between *most* and *more than half* in more detail in Sect. 4.

(i) $[\![\text{most}]\!](A)(B) = 1$ iff $|A \cap B| > |A - B|$

[5] To avoid potentially distracting clutter, we will ignore the difference between sets and pluralities whenever the distinction is immaterial to the discussion. For instance, the symbol for the 'cardinality of' function, | . |, is used for both sets and pluralities.

[6] Throughout this paper, we abstract away from the logical forms corresponding to proportional and superlative truth conditions—see Hackl (2009) as well as Krasikova (2011), Kotek et al. (2011a, b), Szabolcsi (2012), and Pancheva (to appear), among others, for details and discussion—and simply use descriptions of truth conditions such as those in (4b) and (5b) as shorthand.

represent the no-overlap relation, which replaces the non-identity relation $\neq$ of (4b). (5a) is true just in case there is a plurality of students, X, that John talked to that is more numerous than all student pluralities that have no overlap with X. This amounts to demanding that there be a plurality of students that John talked to that is more numerous than the student plurality that John did not talk to. Thus, (5a) expresses proportional truth conditions even though it is analyzed as a superlative construction. It is true just in case John talked to more than half of the students.[7]

The lexical view of *most* and the structural view of *most* make diverging predictions with regard to the ability of bare *most* to take on a superlative reading. Consider a case where bare *most* occurs in subject position, as in (6a). The semantic analyses assumed for such a sentence by the lexical view and the structural view are given in (6b,c), respectively.

(6) a. Most of the dots are blue.
b. $[\![\text{Most of the dots are blue}]\!] = 1$ iff $|\text{Dots} \cap \text{Blue}| > |\text{Dots}| - |\text{Dots} \cap \text{Blue}|$
c. $[\![\text{Most of the dots are blue}]\!] = 1$ iff $\exists d \exists X[\text{Dots}(X) \;\&\; \text{Blue}(X) \;\&\; |X| \geq d \;\&\; \forall Y \in C\, [[\text{Dots}(Y) \;\&\; Y \perp X] \rightarrow |Y| < d]]$

Under the lexical view of *most*, (6a) should only have proportional truth conditions, as given in (7a). Under the structural ambiguity approach to *most*, however, which truth conditions are expressed depends on the content of the comparison class, C. If C is identified with the extension of the plural NP *dots*, i.e. closed under individual sum formation (Link 1983), proportional truth conditions result. This is because all dot pluralities different from the blue dots—whether they are homogenous in color or not—need to be less numerous than the blue dots for the sentence to be true. However, if C is not closed under individual sum formation but further constrained, e.g. so that only homogenously colored dot pluralities are included (as proposed in Kotek et al. 2011a), (6c) only requires that the blue dots outnumber each of the non-blue dot pluralities separately, rather than having to outnumber the non-blue dots as a whole. Under such a construal of C, then, (6a) expresses superlative truth conditions, given in (7b).[8]

(7) a. *Proportional truth conditions*:
$|\text{blue dots}| > |\text{non-blue dots}|$

b. *Superlative truth conditions*:
For each non-blue color Z, $|\text{blue dots}| > |\text{Z dots}|$

To see more concretely how (7a,b) diverge, consider the dot arrays in Figs. 1 and 2. According to the proportional truth conditions, (7a), *Most of the dots are blue* is true

[7] To ensure that (5b) does not express "absolute" truth conditions, which would be paraphrasable by *all the students*, it needs to be assumed that C contains at least two distinct student pluralities. Hackl (2009) argues that this is due to a presupposition of the superlative operator.

[8] Here and throughout, we use 'blue dots' to refer to the target set that is mentioned in the *most* statement and 'yellow dots' and 'red dots' to refer to members of the complement set (the non-blue set). We use the term '*most* statement' to refer to a sentence that has *most* in subject position, e.g. *Most of the dots are blue*, and '*more than half* statement' to refer to a sentence like *More than half of the dots are blue.*

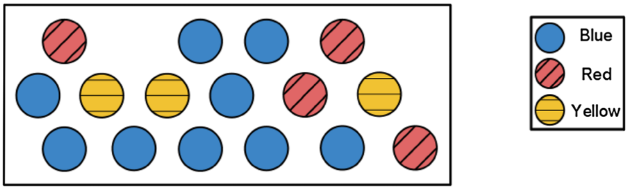

Fig. 1 *Most of the dots are blue* is true under both readings. (Color figure online)

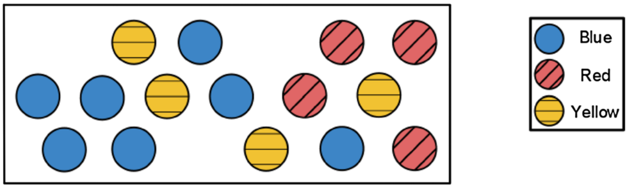

Fig. 2 *Most of the dots are blue* is only true under the superlative reading. (Color figure online)

just in case *more than half* of the dots in the array are blue. An example of such an array is given in Fig. 1, where the 9 blue dots comprise more than half of all the dots and consequently outnumber the non-blue dots. Note that both the lexical view of *most* and the structural view of *most* make the same prediction about the verification of *most* with regard to Fig. 1: speakers should judge the sentence as true in this figure.

The two theories make diverging predictions about the verification of *most* statements with regard to dot arrays as in Fig. 2. This array contains 7 blue dots, 4 red dots and 4 yellow dots. Hence, there are more non-blue dots than blue dots. The lexical view of *most* assigns only proportional truth conditions to the statement *Most of the dots are blue* and therefore predicts that it should be judged false. Under the structural view of *most*, on the other hand, *Most of the dots are blue* has a superlative reading in addition to the proportional reading, and that reading is true in Fig. 2. In particular, under the superlative reading of *most* the number of blue dots is compared to the number of red dots and to the number of yellow dots separately. Since in both comparisons the blue set comes out as more numerous than the competitor, the sentence is true. Therefore, while the lexical view of *most* predicts that speakers will verify the *most* statement as false, the structural view of *most* allows speakers to verify the *most* statement as either true or false with regard to Fig. 2.

In the remainder of this paper we will refer to pictures as in Fig. 2, for which *Most of the dots are blue* is true only under the superlative reading of *most*, as 'superlative' pictures or pictures in the 'superlative' condition. We will use the term 'superlative verification strategy' to refer to the idea that speakers verify a sentence according to superlative truth conditions, and the term 'proportional verification strategy' to refer to the idea that they verify a sentence according to proportional truth conditions.

Two previous studies, Lidz et al. (2011) and Kotek et al. (2011a), have experimentally examined the behavior of *most* in subject position in sentences such as *Most of the dots are blue*, where the sentence is verified against a picture containing blue dots and dots in other colors, in various configurations. Despite many similarities between their studies, they obtained very different results: Kotek et al. (2011a) find that *most* is ambiguous between a dominant proportional reading and a latent superlative reading. The results of Lidz et al.'s (2011) study, on the other hand, are at least consistent with *most* only having a proportional reading.[9] Below, we will present results of a new study that supports the conclusion that bare *most* in subject position does have a superlative reading. This finding provides evidence against the lexical view under which *most* and *the most* are unrelated lexical items. It lends further support to theories such as the one proposed in Hackl (2009), under which *most* and *the most* are both superlative constructions consisting of a silent degree predicate *MANY* and the superlative morpheme *-est*.[10] In Sect. 4.3 we address the source for the divergence of results between Lidz et al. (2011), on the one hand, and Kotek et al. (2011a) and the current study, on the other.

3 Current experiments

As mentioned above, two previous studies of *most*, Lidz et al. (2011) and Kotek et al. (2011a), used similar methodologies to study the behavior of bare *most* in subject position. Both studies tested the verification of *most* statements with respect to dot arrays whose properties were manipulated in various ways. The experiments reported in this section aim to combine the manipulations of the two previous studies to provide conclusive evidence as to the nature of bare *most*.

In particular, we used a Color manipulation of the kind also used in both previous studies, which allowed us to create dot arrays in the 'superlative' condition. We combined this with a Weber ratios manipulation as used in Lidz et al. (2011) but not in Kotek et al. (2011a): under this manipulation, the relative sizes of the blue set and the non-blue set are varied. Verification of *most* statements has been shown by Lidz et al. (2011) to be sensitive to Weber's law, stating that the discriminability of two quantities is a function of their ratio (see Pica et al. 2004; see also Heim et al. 2012, and Tomaszewicz 2011 for relevant work on Polish). Specifically, Lidz et al. (2011) showed that the accuracy of the verification of *most* statements increases gradually as the ratio of blue to non-blue dots in an array increases. Lastly, we added a Determiner manipulation used in Kotek et al. (2011a) but not in Lidz et al. (2011): we compared the verification of *most* to that of *more*

[9] Again, we note that Lidz et al. (2011) do not directly test the existence or absence of the superlative reading, since it is assumed not to exist. Rather, Lidz et al. are concerned with the enumeration of homogeneously colored subsets of the non-blue set—which we will show below is necessary for the calculation of superlative truth conditions—for other reasons. However, leaving the authors' motivations aside, the fact is that the experimental manipulations in Lidz et al.'s study were such that they could have brought out the presence of the superlative reading; yet their results are consistent with no such reading being used by the participants in their experiment.

[10] An anonymous reviewer provides additional evidence in support of the existence of a superlative reading of *most* in subject position, in the form of naturally occurring data such as the following:

(i) Most respondents (34 %) live in the South, followed by 24 % in the Midwest, 23 % in the West, and 17 % in the Northeast.

than half, which is unambiguously proportional. The canonical analysis of *more than half* is given in (8). Under the lexical approach to *most*, given in (6b), *more than half* is truth-conditionally equivalent to bare *most*: both unambiguously have only proportional truth conditions and are true and false in exactly the same circumstances. Under the structural ambiguity approach to *most*, given in (6c), on the other hand, *most* has a reading which *more than half* lacks—the superlative reading—and hence the two determiners are predicted to behave differently under the ‘superlative’ condition: *more than half* will be unambiguously false, but *most* will have a true reading.

(8) $[\![\text{more than half}]\!](A)(B) = 1$ iff $|A \cap B| > \frac{1}{2}|A|$

Finally, the experiments presented here follow Kotek et al. (2011a) in imposing no constraints on how long participants can see the dot array in a trial or how much time they take to make their True/False decision. This diverges from the methodology of Lidz et al. (2011), who only present the dot arrays to their subjects for 150 ms. In Sect. 4.3 we discuss the implications of this choice, which we believe are crucial for understanding the differences between the results obtained by the two previous studies of *most*.

3.1 Experiment 1: ratio manipulation

This experiment provided baseline data on the use of an experimental design that combines the Weber ratios manipulation of Lidz et al. (2011) with the Determiner manipulation of Kotek et al. (2011a). In particular, we were interested in (a) establishing the behavior of *more than half* with regard to dot arrays of varying Weber ratios, and (b) understanding how participants verify *most* and *more than half* when the time allotted to the task is not restricted.

3.1.1 Methods and materials

In each trial in Experiment 1, participants were shown a picture containing 20–21 dots and a sentence describing that picture. Participants were asked to judge whether what the sentence said was ‘True’ or ‘False’ of the picture. Target trials were paired with one of the two statements in (9)–(10), where, as in the Lidz et al. (2011) study, the test sentence was always about the blue dots. Determiner was a between-subject factor: participants saw only *most* statements or only *more than half* statements, and those statements were paired with the exact same pictures across conditions.

(9) Most of the dots are blue
(10) More than half of the dots are blue

All the pictures in Experiment 1 contained blue and yellow dots. Table 1 indicates the number of blue and yellow dots and the blue:yellow ratio for the 9 target trials in Experiment 1. Weber ratios ≤ 1, where the truth conditions of *most* and *more than half* predict that the statements in (9)–(10) are false (‘False-ratios’), are shaded in gray.

Table 1 Blue:yellow dots and ratios in Experiment 1

8:12 [0.67]	9:12 [0.75]	9:11 [0.82]	10:11 [0.91]	10:10 [1]	11:10 [1.1]	11:9 [1.22]	12:9 [1.33]	12:8 [1.5]

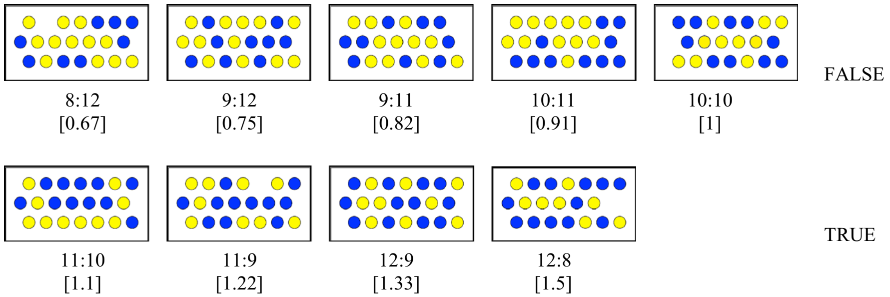

Fig. 3 Dot arrays of target items in Experiment 1. (Color figure online)

Weber ratios > 1, where the truth conditions of *most* and *more than half* predict that the statements in (9)–(10) are true ('True-ratios'), are not shaded.

The 9 target figures used in the experiment are shown in Fig. 3. In addition to the 9 target trials, Experiment 1 contained 24 filler trials. Of these filler items, 8 contained the determiner *more than n* for different numbers *n*, 8 contained the determiner *many*, and 8 contained the determiner *more than n %* or *more than n/m* for different *n*s and *m*s. In half of the filler trials, the correct answer was True and in the other half it was False.[11] The items were presented in one of two pseudo-randomized orders where each pair of target items was separated by at least one filler item, and the first item was not a target item.

The survey, randomization process, and the HTML templates used for this experiment were created using the *turktools* software (Erlewine and Kotek, to appear). The survey was posted on Amazon Mechanical Turk. Participants were paid $0.20 for their participation. They were asked to indicate their native language, but payment was not contingent on their response.

3.1.2 Results

In all, 195 native speakers of English participated in this study.[12] Of these, 97 subjects participated in the 'most' condition and 98 participated in the 'more than half'

[11] The pictures associated with *many* and with the proportional determiners were generated so as to clearly match the expressed truth conditions in the True condition and to clearly not match them in the False condition. That this manipulation was successful was verified by a post-hoc inspection of accuracy of the filler items, detailed in Sect. 3.2.2 below.

[12] Here and in Experiment 2 results are reported for all native speakers of English, including those who speak a second language. The results do not change if only monolingual speakers are included in the analysis.

condition. Four non-native speakers and one subject who did not report on their native language were excluded from the analysis. No subjects were excluded from the analysis because of low accuracy rates (<75 % on filler trials). Two filler items were excluded from the analysis because of low accuracy rates (<75 % accuracy across all participants). The remaining 22 filler items had a mean accuracy of 97.3 %. Our results would not change if a threshold of 80 % had been chosen instead.

Figure 4 shows the average percentage of True responses to *Most of the dots are blue* and *More than half of the dots are blue* for the 9 ratios of Experiment 1 ($N = 195$). We observe an inflection point: for False-ratios, the percentage of True responses is near zero. For True-ratios, the percentage of True responses is at 77 % and 85 % for *most* and *more than half* respectively for the ratio 1.1, and at 90 % and nearly 100 % for *most* and *more than half* respectively for all ratios above 1.1.

Importantly, we also observe an asymmetry between the behavior of *most* and *more than half*: although the verification behavior of *most* is almost identical to that of *more than half* for all False-ratios, the two determiners come apart for True-ratios. We observe a parallel proportion of True judgments for *most* and *more than half*, with *most* consistently verified as true 5–7 % less often than *more than half*.

A mixed-effects logit model was fit to the data.[13] The model examines the effect of the Weber ratios (with 9 levels as specified in Table 1) and Determiner (*most* vs. *more than half*) on percent-True in Experiment 1. The random effect structure includes random intercepts for both subjects and items, and by-subject random

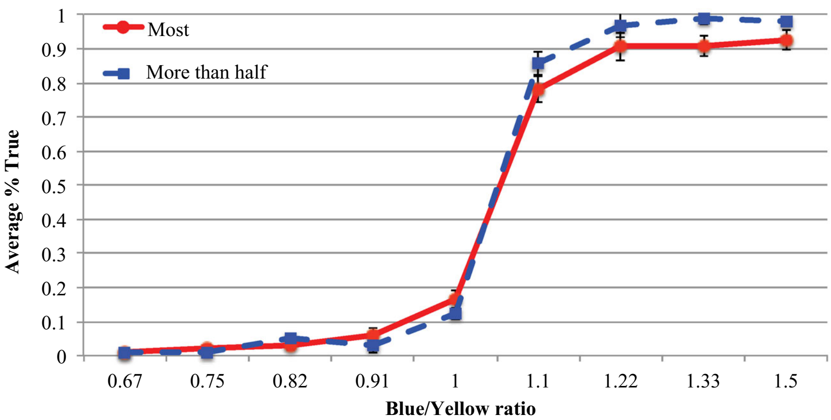

Fig. 4 Average percentage of True responses in Experiment 1 (accuracy ≥75 %). (Color figure online)

[13] The models reported in the paper were fit using R and the R package lme4 (Bates and Maechler 2009). The Determiner predictor was contrast coded as follows:

(i) Determiner: *most* = 0.5 *more than half* = −0.5

Random effect structures in our models are the maximal ones supported by the data and by log-likelihood tests comparing models with the effects to models from which they were removed (cf. Baayen et al. 2008; Barr et al. 2013).

Table 2 Summary of the fixed effects

Predictor	Coefficient	Standard error	p
Intercept	−20.599	1.600	**<0.001**
Determiner	4.923	2.597	0.0580
Weber ratios	19.216	1.516	**<0.001**
Determiner × Weber ratios	−4.968	2.487	**<0.05**

Bold values are statistically significant

slopes for the effects of Weber ratios.[14] The fixed effects are summarized in Table 2. The correlations among the fixed effects are all within ±0.3, with the exception of a strong correlation between the main effect of Determiner and its interaction with Weber ratios.[15]

We find a main effect of Weber ratios and an interaction between Weber ratios and Determiner. Because of the strong co-linearity between the main effect of Determiner and the interaction of Determiner and Weber ratios, we do not put any weight on the near-significant main effect of Determiner found in this model.

3.1.3 Discussion

Experiment 1 yielded two important findings. First, the verification of *most* and *more than half* is indeed sensitive to the Weber ratio, such that for arrays with ratios below 1, the test sentences are judged false and for arrays above 1 they are judged true. This is, of course, just a reflection of the truth-conditional import of these determiners. Second, we find that although *most* and *more than half* are truth-conditionally equivalent in all target items of Experiment 1, the verification of the two determiners is notably different: *most* patterns with *more than half* for verification of items with ratios below and equal to 1: for these, the test sentences are judged to be true less than 10 % of the time; for ratios above 1, on the other hand, *most* and *more than half* exhibit parallel behavior, but the test sentence with *most* is judged to be true on average 5–7 % less often than that with *more than half* in all 4 True items. We return to this point in the discussion in Sect. 4.1. To conclude, although the task demands are different in Experiment 1 compared to the task in Lidz et al. (2011), we find a similar effect of the Weber ratios on the

[14] A more detailed model that includes slopes for the random effect of Determiner does not significantly improve the model fit compared to the model that does not include these slopes ($\chi^2(3) = 4.7976$, $p = 0.1872$).

[15] A model that predicts percent-True in Experiment 1 from Determiner and Truth conditions (predicting False for all Weber ratios, ≤ 1 and True for all Weber ratios > 1 for both *most* and *more than half*) yields main effects of these two predictors and no interaction. A model that predicts percent-True in Experiment 1 from Determiner, Weber ratios, and Truth conditions did not converge because of strong co-linearity between the latter two predictors. The model that uses Weber ratios as a predictor is more informative than the model that uses Truth conditions, given the design of Experiment 1. This model allows for pair-wise comparisons of the behavior of participants at different ratios, as opposed to collapsing all different points and testing one cloud of noisier data. We therefore concentrate on this model, rather than on the one that uses Truth conditions in place of Weber ratios.

verification of *most*. Experiment 2 builds on this finding by adding a Color manipulation to the Weber ratios manipulation.

3.2 Experiment 2: ratio-by-color manipulation

Experiment 2 expands on the results of Experiment 1 and explores the simultaneous effect of Weber ratios and Color on the verification of statements containing *most* and *more than half*. Pictures in this experiment have either two (blue and yellow) or three (blue, yellow, and red) colors. The presence of a third color in the picture is predicted to have an effect only on those determiners that can be evaluated using subsets of the non-blue set separately. Consequently, *more than half* is expected not to be sensitive to the Color manipulation. Under the lexical view of *most*, *most* too is predicted not to be sensitive to the Color manipulation, because it only has a proportional reading. Under the structural view, however, *most* is predicted to be sensitive to the Color manipulation, because it has a superlative reading under which the blue set is compared to the yellow set and to the red set separately. Moreover, a 'superlative' verification strategy of *most* is expected to be sensitive not only to the presence of multiple subsets of non-blue dots, but also to their structure: following Weber's law, a *most* statement is predicted to be more difficult to verify against a picture with a blue:yellow:red ratio of 10:10:1, compared to a picture with a ratio of 10:6:5, even though the overall blue:non-blue ratio is the same in both cases. To investigate this prediction, Experiment 2 tests not only the effect of Color but also the effect of different Weber ratios within the non-blue set on the verification of *most* and *more than half*.

3.2.1 Methods and materials

Experiment 2 combines the Determiner and Weber ratios manipulations of Experiment 1 with an additional Color manipulation: dot arrays in Experiment 2 had either two colors or three colors. We used three different ways of constructing 3-color arrays by varying the ratios of blue:yellow and blue:red dots. In what follows, we will refer to these ratios as Color ratios. Pictures in the 2-Color condition contained blue and yellow dots, and pictures in the 3-Color condition contained blue, yellow, and red dots. In the balanced condition, the yellow and red dots were split up evenly. In the mildly balanced condition, the dots were split up somewhat evenly, but with more yellow than red dots. In the unbalanced condition, pictures contained one red dot and all the other non-blue dots were yellow. Pictures used 9 different Weber ratios: four below 1; one at exactly 1; and four above 1; all were identical to those used in Experiment 1. These Weber ratios were held constant across the 2-Color condition and the three 3-Color conditions. The overall number of dots in the pictures was again 20–21.

The design of Experiment 2 is summarized in Table 3. Each row corresponds to one Color level (2C, 3C unbalanced, 3C mildly balanced, 3C balanced). Each column represents one Weber ratio, indicated at the top of the column. The Color ratio is manipulated in the different Color conditions, with a total of 26 different

Table 3 Blue:yellow:red dots, and Color ratios (in brackets) in Experiment 2

TC / COLOR	Proportional *most* : False 8:12	9:12	9:11	10:11	10:10	Proportional *most* : True 11:10	11:9	12:9	12:8
2C	8:12 [0.67]	9:12 [0.75]	9:11 [0.82]	10:11 [0.91]	10:10 [1]	11:10 [1.1]	11:9 [1.22]	12:9 [1.33]	12:8 [1.5]
3C UNBALANCED	8:11:1 [0.73]	9:11:1 [0.86]	9:10:1 [0.9]	10:10:1 [1]	10:9:1 [1.11]	11:9:1 [1.22]	11:8:1 [1.38]	12:8:1 [1.5]	12:7:1 [1.71]
3C MILDLY BALANCED	8:9:3 [0.89]	9:9:3 [1]	9:8:3 [1.13]	10:8:3 [1.25]	10:7:3 [1.42]	11:7:3 [1.57]	11:6:3 [1.83]	12:7:2 [1.71]	12:6:2 [2]
3C BALANCED	8:6:6 [1.33]	9:6:6 [1.5]	9:6:5 [1.5]	10:6:5 [1.67]	10:6:4 [1.67]	11:5:5 [2.2]	11:5:4 [2.2]	12:5:4 [2.4]	12:4:4 [3]

Color ratios distributed across the Color conditions. Each cell in Table 3 provides information about the numbers of blue, yellow, and red dots in that cell, and the Color ratio of blue:yellow dots in the cell (in brackets). Note that for cells in the 2C condition, the Color ratio is the same as the Weber ratio, since there is only one non-blue color in those trials.

In Table 3 there are four columns on the right with Weber ratios above 1, for which *most* statements are true under proportional (and hence also superlative) truth conditions. In cells in the first five columns on the left, *most* and *more than half* statements are false under proportional truth conditions. These cells are divided into two groups, as indicated by their shading: the gray shaded cells represent pictures for which *most* is also false under superlative truth conditions. The white cells in the table represent 'superlative' pictures, which are false under proportional truth conditions but true under superlative truth conditions. These pictures contain dot arrays with a Weber ratio below 1 (as can be seen in the corresponding 2C cell in the same column), but a Color ratio above 1, as the numbers in brackets in those cells show.

The 36 target pictures used in Experiment 2 are given in Fig. 5. Below each picture we indicate the numbers of blue:yellow:red dots in the picture and the blue: yellow Color ratio (in brackets) for that picture. For the 2C items, the same pictures were used as in Experiment 1.

We can make the following prediction: if *most* has a superlative reading—that is, if *most* can be verified using the Color ratios (blue:yellow and blue:red) instead of the Weber ratio (blue:non-blue), we expect speakers to judge *most* statements for 'superlative' pictures as true more often than for other pictures in the same column (i.e., pictures that have the same Weber ratio but are false under the superlative reading). Furthermore, this behavior should track Weber's law similarly to the behavior we observed for Weber ratios in Experiment 1: higher Color ratios should be easier to verify than ratios that are closer to 1. On the other hand, if *most* only has a proportional reading, which is truth-conditionally equivalent to that of *more than half*, it should not be affected by Color ratios. In that case, we predict all cells in a given column in Table 3 to be similarly rated for *most*, and the same is

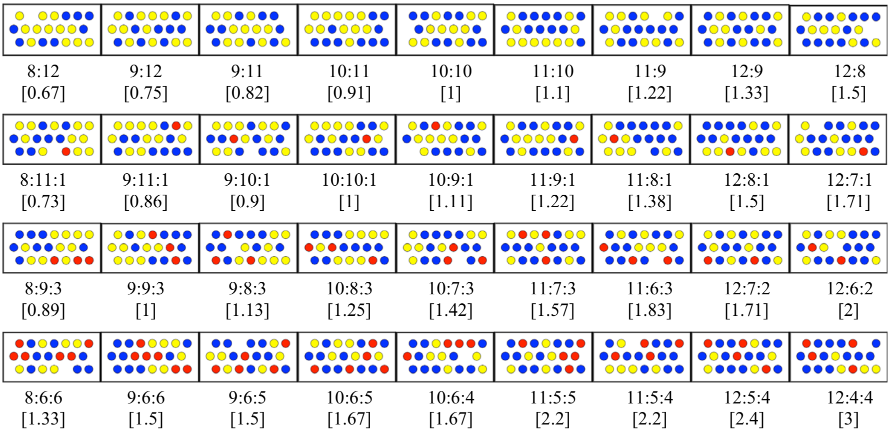

Fig. 5 Dot arrays of target items in Experiment 2. (Color figure online)

expected for *more than half* as well. The predictions for the trials that are false under the proportional reading are summarized in (11) below.[16]

(11) Predictions for Experiment 2 (5 columns on the left, where proportional most is false):
- a. *More than half*: No effect of Color.
  All trials in columns 1–5 are judged False.
- b. *Most*—lexical view: No effect of Color.
  All trials in columns 1–5 are judged False.
- c. *Most*—structural view: Effect of Color.
  2C: All trials in columns 1–5 are judged False.
  3C unbalanced: Trials in columns 1–4 are False; the trial in column 5 is True.
  3C mildly balanced: Trials in columns 1–2 are False; trials in columns 3–5 are True.
  3C balanced: All trials in columns 1–5 are True.

To test these predictions, target trials in Experiment 2 were paired with one of the two statements in (12)–(13) about the blue dots. As in Experiment 1, Determiner was a between-subject factor: participants evaluated the pictures shown in Fig. 5 either against *most* statements or against *more than half* statements (but no participant was given both *most* and *more than half* statements).

(12) Most of the dots are blue
(13) More than half of the dots are blue

[16] Both theories of *most* predict that trials rendering the test sentences true under the proportional reading will indeed yield True judgments from speakers: the superlative reading is entailed whenever proportional truth conditions are satisfied.

The experiment contained 72 filler items, in addition to the 36 target items. Of the filler items, 24 contained the determiner *more than n* for different *n*s, 24 contained the determiner *many*, and 24 contained the determiner *more than n %* or *more than n/m* for different *n*s and *m*s. In half of the filler trials, the correct answer was True and in the other half it was False.

As before, the items were presented in surveys in one of 8 pseudo-randomized orders where each pair of target items was separated by at least one filler item, and the first item was not a target item. The survey, randomization process, and the HTML templates used for this experiment were created using the *turktools* software (Erlewine and Kotek, to appear). The surveys were posted on Amazon Mechanical Turk. Participants were paid $0.20 for their participation. They were asked to indicate their native language, but payment was not contingent on their response.

3.2.2 Results

Data from 251 native speakers of English who participated in this study was included in the analysis. Of those, 135 subjects participated in the 'most' condition and 116 participated in the 'more than half' condition. Three participants were excluded from the analysis because of low accuracy rates (<75 % for the *more than n* trials);[17] nine were excluded because they were non-native speakers; and 62 participants were excluded because they did not complete one or more trials.[18,19]

Figure 6 shows the average percentage of True responses to *most* statements on the left and to *more than half* statements on the right for the 9 Weber ratios of Experiment 2. Each Color condition is plotted separately. For all four Color conditions, we observe a clear inflection point: for Weber ratios >1—where the truth conditions of *most* and *more than half* predict that the statements in (12)–(13) are true—the percentage of True responses is above 65 % for the ratio 1.1 and near 90 % for the other data points. For Weber ratios ≤ 1—where the proportional truth conditions of *most* and *more than half* predict that the corresponding *most* and *more than half* statements are false—the percentage of True responses is very low.

Importantly, we also see in Fig. 6 an effect of the Color manipulation in the 'superlative' trials for *most* but not for *more than half* such that *most* yielded more True judgments in the conditions with Weber ratios ≤ 1. To discuss this in more detail, let us examine the four graphs in Fig. 7. These graphs compare participants'

[17] Three additional subjects would be excluded if accuracy were calculated based on all of the fillers. Here we chose to exclude only based on *more than n* items because these had the same blue:non-blue ratios as the target items.

[18] As an anonymous reviewer points out, this is quite a large number of participants to exclude, but given the fact that there were 108 trials in total and that they were all presented on a single page on a computer screen one after another, it is not at all surprising that many participants inadvertently skipped one or more trials.

[19] Two additional participants would be excluded if an 80 % accuracy rate were used. In addition, if accuracy on all filler items were used instead of accuracy only on *more than n* items, a total of 6 additional participants would be excluded (that is, 12 instead of 6, which are excluded with the 75 % rate). None of these choices affect the statistical findings that we report below: they remain intact under all of these possible exclusion criteria.

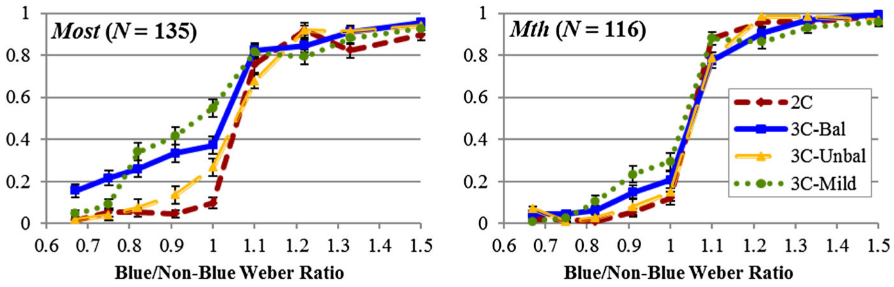

Fig. 6 Percent-True for *most* (*left*) and *more than half* (*Mth*, *right*) in Experiment 2. (Color figure online)

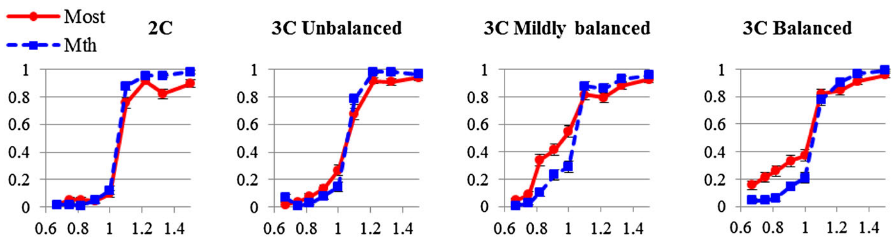

Fig. 7 Comparison of *most* and *more than half* (*Mth*) broken down by COLOR conditions

behavior for each Color condition for *most* and *more than half*. Recall that Determiner was a between-subject factor and that participants saw the same pictures matched with a *most* statement in the 'most' condition and with a *more than half* statement in the 'more than half' condition. Several effects of interest are apparent in these graphs.

Zooming in on the behavior of *most* for Weber ratios ≤ 1 in the different Color conditions, we observe a difference in the behavior of *most* compared to *more than half*. Although all pictures have Weber ratios ≤ 1 (false under proportional truth conditions), some pictures have Color ratios >1 (true under superlative truth conditions). These pictures are: (a) in the unbalanced condition, data point 5 (Weber ratio = 1; Color ratio = 1.11); (b) in the mildly balanced condition, data points 3–5 (Weber ratios = 0.82, 0.91, 1; Color ratios = 1.13, 1.25, 1.42); and (c) in the balanced condition, data points 1–5 (Weber ratios = 0.65, 0.75, 0.82, 0.91, 1; Color ratios = 1.33, 1.5, 1.5, 1.67, 1.67). As can be observed in Figs. 6 and 7, we see an increase in True responses to *most* statements compared to *more than half* statements precisely in these 9 pictures.[20]

[20] One additional finding is an unexpectedly high rate of True answers to the 'false'-trials in the 3C mildly balanced condition and some trials in the 3C unbalanced condition. In Sect. 3.3 we will argue that this is a spurious finding, which is the result of noisy pictures in those trials and is unrelated to the experimental manipulations in Experiment 2.

Turning now to the 2C graph, which contains data from the same pictures that were used in Experiment 1 (see Fig. 4), we can again clearly observe the behavioral asymmetry of *most*: *most* behaves the same as *more than half* for Weber ratios ≤ 1: it is judged True at very low rates, close to 0 %. For Weber ratios >1, however, the behavior of *most* diverges from that of *more than half*. *More than half* is symmetric and sharply changes from True rates close to 0 % to True rates close to 100 % precisely when its truth conditions predict that it would: at the first Weber ratio >1. *Most*, on the other hand, has a more gradual increase in True judgments for Weber ratios >1, and even at the highest Weber ratio, 1.5, it is still judged True less often than *more than half*. In fact, as the graphs in Fig. 7 indicate, in Weber ratios >1 for all COLOR conditions, *most* is consistently judged True less often than *more than half* for the True Weber ratios.[21] This is the same pattern that we observed in Experiment 1. We will expand on this point in Sect. 3.3.

A mixed-effects logit model was fit to the data. The model examines the effect of WEBER RATIOS (with 9 levels as specified in Table 3 above), COLOR RATIOS (with 26 levels as specified in Table 3 above), and DETERMINER (*most* vs. *more than half*) on percent-True. The random effect structure includes random intercepts for both subjects and items, and by-subject random slopes for the effects of WEBER RATIOS.[22] The fixed effects are summarized in Table 4. The correlations among the fixed effects are all within ±0.4, with the exception of a strong correlation between the main effect of DETERMINER and the main effect of WEBER RATIOS and the interaction of these two factors.

The results show main effects of DETERMINER and of WEBER RATIOS. That is, we find that *most* is verified differently than *more than half*, but they are both affected by the WEBER RATIOS manipulation. Additionally, we find a DETERMINER × WEBER RATIOS interaction, such that *more than half* is affected by the WEBER RATIOS manipulation more than *most* is, and a DETERMINER × COLOR RATIOS interaction, such that *most* is affected by the COLOR RATIOS manipulation more than *more than half* is.

Table 4 Summary of the fixed effects in Experiment 2

Predictor	Coefficient	SE	z value	*p*
Intercept	−22.1089	1.3689	−16.151	<**0.001**
DETERMINER	8.2047	1.3492	6.081	<**0.001**
WEBER RATIOS	20.1878	1.4169	14.248	<**0.001**
COLOR RATIOS	0.5504	0.4510	1.220	0.222
DETERMINER × WEBER RATIOS	−9.4825	1.3114	−7.231	<**0.001**
DETERMINER × COLOR RATIOS	1.5190	0.2893	5.251	<**0.001**

Bold values are statistically significant

21 An inspection of individual subjects reveals that this behavior cannot be attributed to some small subset of participants who consistently judged *most* trials in our 'true'-conditions as False.

22 A model that also included the ratio of blue to red dots in the array did not converge (recall that COLOR RATIOS indicate the blue to red ratio in our pictures). A model that also includes superlative and proportional truth conditions reveals similar effects to the ones reported above, but suffers from a high degree of colinearity because of the inclusion of the WEBER RATIOS and COLOR RATIOS predictions.

3.2.3 Discussion

We see two main effects in Experiment 2: Weber ratios and Determiner. We find that Weber ratios affect the verification of both *most* and *more than half* such that the True responses to both *most* and *more than half* statements increase as the ratio of blue to non-blue increases. The main effect of Determiner indicates that *most* is judged True more often than *more than half* is. As can be seen in Fig. 6, this result is caused by the 'superlative' pictures in the 'most' condition, where we observe that *most* statements are judged True more often than *more than half* statements are. In fact, as this observe suggests, the main effect of Determiner is driven by the Determiner × Color ratios interaction—that is, we see an increase in True responses to 'superlative' pictures for *most* but not *more than half*. The Determiner × Weber ratios interaction suggests that Weber ratios contribute more to the verification of *more than half* than to the verification of *most*. This is so because, as we have seen, more factors contribute to the verification of *most* statements than of *more than half* statements, and these factors mediate the contribution of the Weber ratios themselves.

One final result of Experiment 2 is the replication of Experiment 1 in the 2-Color condition: we again observe the asymmetric behavior of *most*, where 'false' trials are judged as False close to 100 % of the time, but 'true' trials are judged as True only 90 % of the time or less. For *more than half*, by comparison, we find no difference between the verification of False and True trials: both are at close to 100 %. Furthermore, as in Experiment 1, we see that trials with Weber ratio = 1 are verified as False close to 90 % of the time. These results suggest that although participants in Experiment 2 were exposed to many more pictures than participants in Experiment 1, the diversity of the pictures and the increase in the number of decisions required did not change the nature of the judgments that we observed in the two experiments. That is, we find no evidence for a difference in participants' strategies across the two experiments.

3.3 Implications: *most* in subject position has a genuine superlative reading

In this section we discuss the results of Experiments 1 and 2 with regard to the debate over the correct analysis of *most*. We will show that postulating a superlative reading of bare *most* in subject position explains the pattern of results observed in the experiments, and we argue that these results are not predicted under the lexical view.

Recall that the starting point for our experiments was the comparison between a lexical and a structural view of *most*. The lexical view predicts that bare *most* only has proportional truth conditions, while the structural view predicts that bare *most* is ambiguous between the two readings in (14), repeated from (7) above.

(14) a. Most of the dots are blue

b. *Proportional truth conditions:*
$|\text{blue dots}| > |\text{non-blue dots}|$

c. *Superlative truth conditions:*
For each non-blue color Z, $|\text{blue dots}| > |\text{Z dots}|$

For the superlative reading to be detectable, dot arrays in the 'superlative' condition are required. That is, we need pictures that make a *most* statement false under proportional truth conditions but true under superlative truth conditions. If speakers verify *most* statements only according to proportional truth conditions, we expect them to judge *Most of the dots are blue* as false for such pictures. If they can access superlative truth conditions, they will be able to judge the same statement as true for those pictures. All speakers are expected to judge a *more than half* statement as false when verified against those same pictures.

Previous experimental works (Lidz et al. 2011; Tomaszewicz 2011; Heim et al. 2012) have shown that *most* is sensitive to Weber's law such that *most* statements are more difficult to verify in arrays with close blue:non-blue ratios compared to arrays in which the two numerosities are further apart. Experiment 1 showed that this result holds not only when the arrays are presented for very short durations but also when participants are given as much time as they need to make their decision.

Experiment 2 built on the Weber ratios manipulation of Experiment 1 and compared subjects' behavior with respect to 'superlative' pictures in which not only the blue:non-blue ratio was varied, but also the composition of the non-blue set. We tested three different 'superlative' conditions, where the ratios of blue:yellow and blue:red dots (Color ratios) were systematically varied. Following Weber's law, we suspected that if participants use a 'superlative' verification strategy to verify *most* statements, the ease with which they use this strategy will be affected by the ratios of the blue:yellow and blue:red comparisons. If the ratios are closer to 1, the verification of a *most* statement will be difficult; on the other hand, as the ratios get larger, speakers will have an easier time verifying the statement as true under superlative truth conditions.

Indeed, Experiment 2 showed that when speakers verify *most* statements, 'superlative' pictures were judged true more often than other pictures with the same blue:non-blue ratio that were false under the superlative reading. Furthermore, there was an increase in the rate of True answers to *most* statements as the Color ratios increased. This pattern is consistent with the view that bare *most* has a superlative reading and it is inconsistent with the view that bare *most* only has a proportional reading that is truth-conditionally equivalent to *more than half*.

Upon closer examination of the 'superlative' pictures, we note that although the rates of True responses to *most* statements were clearly above those for other pictures with the same ratios that were not in the 'superlative' condition, these rates were rather low overall and, in fact, almost without exception below 50 %. This seems prima facie unexpected on the structural view, since this view predicts that superlative pictures can be judged as true under the superlative reading of *most*. One might thus expect substantial rates of True responses to 'superlative' pictures, reflecting the superlative reading of *most*. However, this prediction can only be made if the two readings of *most* are equally accessible to speakers during the verification task. As was shown by Kotek et al. (2011a), the ambiguity of bare *most* in subject position is in fact heavily biased in favor of the proportional reading. That is, the superlative reading of *most* is latent and often masked by the more dominant proportional reading. Moreover, Kotek et al. (2011a) showed that the superlative reading of *most* was only available to about one third of the speakers in each of their three experiments. Assuming the same prevalence of the superlative reading in our participants, this means that the rates of True responses to 'superlative' pictures should be lower than the rates of True responses to pictures that are true on both the proportional and the superlative reading. In the latter case a True response is available to all speakers, under whichever reading they verify, while in the former a True response is available only to some of the speakers some of the time.

To further investigate the hypothesis that the superlative reading was only accessible to some of the participants in Experiment 2, below we classify participants in the 'most' condition in Experiment 2 as "proportional" or "superlative" speakers according to whether or not they accessed superlative truth conditions in the verification of *most*.[23] The criterion we use for this classification of participants is given in (15).

(15) "*Superlative*" vs. "*proportional*" classification of participants in Experiment 2:

In proportionally-false pictures: if percent-True in 3C BALANCED > percent-True in 3C UNBALANCED, then *superlative*; otherwise, *proportional.*

The idea is that if speakers do not use superlative truth conditions in the verification of bare *most*, they will verify all proportionally-false trials as false with the same error rate: that is, they are not more likely to reply True to pictures in the 'superlative' condition than to other pictures with the same Weber ratio that are false under the superlative reading. On the other hand, speakers who use a superlative verification strategy are more likely to verify a *most* statement as true in 'superlative' pictures compared to other pictures with the same Weber ratio that are false under superlative truth conditions. Hence, if speakers replied True more often in the 3C BALANCED condition (where all five proportionally-false trials could be judged as true under superlative truth conditions) than in the 3C UNBALANCED condition (where only one proportionally-false trial could be judged as true under superlative truth conditions), we classify them as "superlative"; otherwise we classify them as "proportional" speakers.[24]

Using the criterion in (15), we find that 79 participants in the *most* condition are classified as "proportional" speakers, and the other 56 are classified as "superlative." These numbers are comparable with Kotek et al.'s (2011a) numbers. Figure 8 shows the percentage of True responses to all target trials in Experiment 2 broken down by the classification of "proportional" vs. "superlative" *most* speakers (*most-prop* vs. *most-sup*).

[23] Kotek et al. (2011a) suggest that speakers for whom bare *most* is ambiguous between a proportional and a superlative reading use both truth conditions in the verification of *most* statements in the same study, rather than only using superlative truth conditions (speakers who consistently only used proportional truth conditions in the verification of *most* were classified as "proportional" speakers). Therefore, our classification refers to speakers' ability to optionally use a 'superlative' verification strategy, rather than singling out speakers who *only* use such a strategy.

[24] We do not use the 2C condition in the classification, as it is likely to contain less noise than the 3C conditions due to the relative simplicity of the task, and hence would make the criterion more inclusive. Also, as the reader will see below, we will use the data from the 3C MILDLY BALANCED condition, which we do not directly manipulate here, in our analysis.

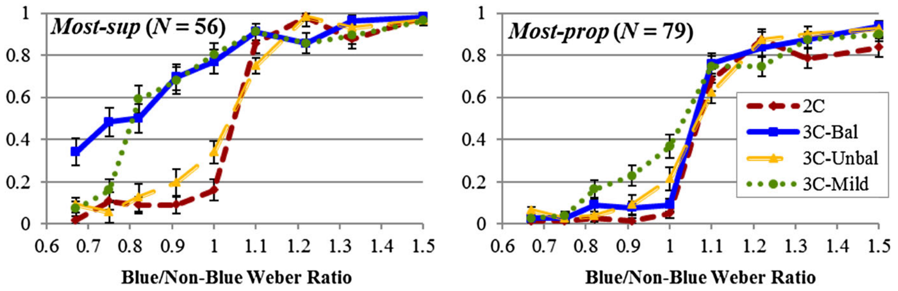

Fig. 8 Percent-True for "superlative" (*most-sup*, *left*) vs. "proportional" (*most-prop*, *right*) *most* speakers in Experiment 2. (Color figure online)

The classification of speakers in Fig. 8 indeed sharpens the results: we find that there is little variation among the different COLOR conditions for "proportional" participants, but there is a strong effect within the "superlative" participants. In particular, we see that in the case of *most* verified by "superlative" participants (Fig. 8, left panel) all five 'superlative' trials in the 3C BALANCED condition, three trials in the 3C MILDLY BALANCED condition, and the one trial in the 3C UNBALANCED condition are judged true more often than in the case of *most* verified by "proportional" participants (Fig. 8, right panel). This mirrors the design of the items in Table 3, repeated below for convenience. Specifically, the items that exhibit an increase in True-responses are those that are described in the white region on the left side of Table 3. They are the items that use pictures in the 'superlative' condition: they are true under a superlative reading as soon as the number of blue dots is greater than the number of yellow and red dots separately (the superlative reading), but false under a proportional reading.

To see whether the classification in (15) identifies internally consistent subgroups of speakers in Experiment 2, we can compare the behavior of participants who were classified as "proportional" and the behavior of participants in the 'more than half'

Table 3 (Repeated): Blue:yellow:red dots, and COLOR RATIOS (in brackets) in Experiment 2

TC / COLOR	Proportional *most* : False					Proportional *most* : True			
	8:12	9:12	9:11	10:11	10:10	11:10	11:9	12:9	12:8
2C	8:12 [0.67]	9:12 [0.75]	9:11 [0.82]	10:11 [0.91]	10:10 [1]	11:10 [1.1]	11:9 [1.22]	12:9 [1.33]	12:8 [1.5]
3C UNBALANCED	8:11:1 [0.73]	9:11:1 [0.86]	9:10:1 [0.9]	10:10:1 [1]	10:9:1 [1.11]	11:9:1 [1.22]	11:8:1 [1.38]	12:8:1 [1.5]	12:7:1 [1.71]
3C MILDLY BALANCED	8:9:3 [0.89]	9:9:3 [1]	9:8:3 [1.13]	10:8:3 [1.25]	10:7:3 [1.42]	11:7:3 [1.57]	11:6:3 [1.83]	12:7:2 [1.71]	12:6:2 [2]
3C BALANCED	8:6:6 [1.33]	9:6:6 [1.5]	9:6:5 [1.5]	10:6:5 [1.67]	10:6:4 [1.67]	11:5:5 [2.2]	11:5:4 [2.2]	12:5:4 [2.4]	12:4:4 [3]

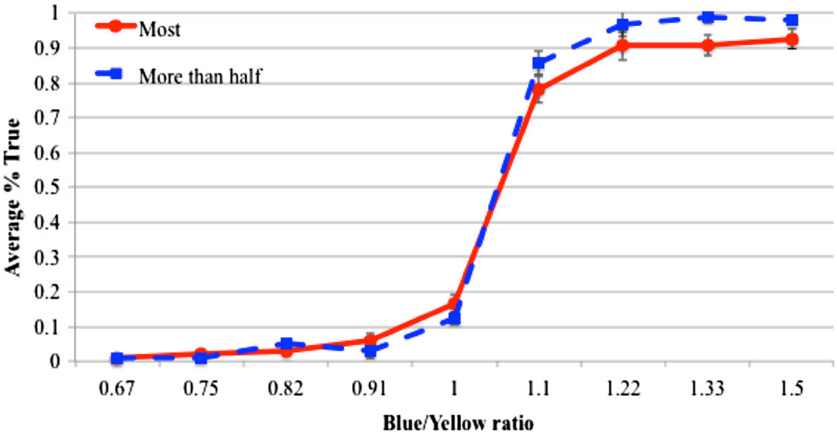

Fig. 4 (repeated): Average percentage of True responses in Experiment 1. (Color figure online)

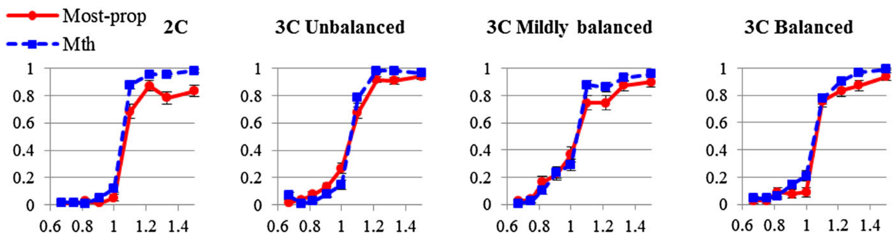

Fig. 9 Comparison of "proportional" *most* speakers (*most-prop*) and *more than half* (*Mth*) broken down by COLOR conditions. (Color figure online)

condition. Since both groups of speakers are assumed to have verified *most* and *more than half* using a proportional verification strategy exclusively, and since *most* and *more than half* are truth-conditionally equivalent for those speakers, we expect them to show similar verification behavior for *most* and *more than half*. More specifically, we expect similar results to those found in Experiment 1; the relevant figure (Fig. 4) is repeated here for convenience.

Since *most* is not sensitive to COLOR under the proportional reading, we expect to see similar behavior across all four COLOR conditions of Experiment 2 for *most* and *more than half*: we expect all false items to be verified as false at very high rates for both *most* and *more than half*; we expect *more than half* to be verified as true at high rates for all true items, and we expect *most* to be verified as true less often than *more than half* for those same trials (in light of the behavioral asymmetry of *most*). Figure 9 shows the comparison of "proportional" *most* speakers (*most-prop*) and *more than half*, broken down by COLOR conditions.

Several effects can be observed in the graphs in Fig. 9. First, we observe that the behavior of *most* largely tracks the behavior of *more than half* for proportionally-false trials. This is as expected if the speakers we classified as "proportional" can access only a proportional reading for *most.* Moreover, we see that in proportionally-true trials, *most* is consistently verified as true less often than *more than half* is. *More than half* is generally verified as true at close to ceiling rates for all 'true' trials, but *most* is almost never verified at the same rates even at the highest Weber ratio of 1.5. Finally, we observe that whatever noise is introduced by the particular items used in Experiment 2 (see Figs. 5 and 6) affects *most* and *more than half* to the same extent. This noise is particularly visible in the 'false' trials of the 3C MILDLY BALANCED condition and some 'false' trials in the 3C UNBALANCED condition, but importantly it does not have a greater effect on one determiner than the other. This justifies the comparisons we have made here between *most* and *more than half*; furthermore, it motivates close scrutiny of deviations in the verification of *most* by "superlative" speakers that go beyond the baseline we have just seen.

Next let us compare the behavior of participants who were classified as "superlative" with the behavior of participants who were classified as "proportional" in the 'most' condition. Figure 10 shows the comparison of *most-sup* and *most-prop*, broken down by COLOR conditions.

Figure 10 clearly demonstrates the presence of a superlative reading of *most* for participants who were classified as "superlative" speakers. In general, we observe that *most* statements are verified as true more often by "superlative" speakers than by "proportional" speakers for all Weber ratios in all COLOR conditions. This property of the "superlative" speakers is particularly pronounced in the 'superlative' pictures, where the rates of True responses to *most* statements are clearly above those of parallel trials for "proportional" speakers, whose responses, in turn, resemble rates of True responses observed earlier for *more than half* statements. Furthermore, we observe that the rate of True responses to 'superlative' pictures increases as the Color ratios increase, as expected based on the results of Experiment 1 as well as the results of Lidz et al. (2011).

This finding is precisely what is predicted if bare *most* in subject position has a latent superlative reading that is accessible to some speakers (that is, the "superlative" speakers), but it is unexpected if *most* does not have such a reading. If such a reading did not exist, we would expect to only see behavior consistent with the proportional reading of *most*, but not the superlative reading of *most*. However, we see a clear difference in the judgment patterns of "superlative" and "proportional" speakers precisely in those pictures that correspond to the

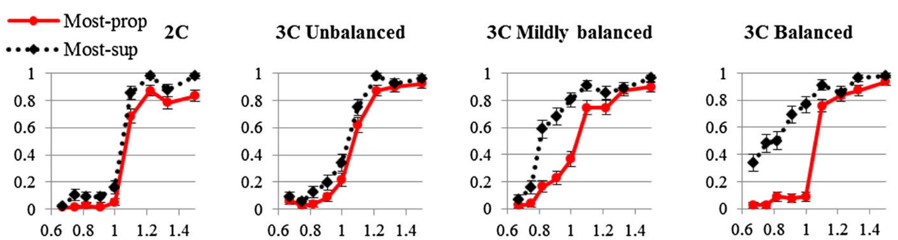

Fig. 10 Comparison of "superlative" (*most-sup*) and "proportional" (*most-prop*) *most* speakers broken down by COLOR conditions

'superlative' condition—that is, pictures that can be judged as true under superlative truth conditions even though they are false under proportional truth conditions.

It is particularly illuminating to compare the behavior of the two speaker groups in the 3C MILDLY BALANCED condition: this condition did not enter into the calculation used to classify speakers into the two groups in (15), and hence represents the effect of this classification independent of the data used in the classification itself. We observe that the four (proportionally and superlatively) 'true' ratios in this condition were verified as true with similar rates by "proportional" and "superlative" speakers. Moreover, we can clearly observe that the first two ratios on the left (false under both readings of *most*) were verified as false at similar rates for both speaker groups, but the next three ratios, all 'superlative' conditions, exhibit a much higher rate of True responses in the "superlative" group than in the "proportional" group. Recall that, as we observed in Fig. 9, the increase in True rates for these pictures in "proportional" speakers is parallel to that observed for *more than half* and hence likely reflects noise present in our items, rather than superlative verification behavior. The increase observed beyond that for "superlative" pictures cannot be similarly attributed to noise and must correspond to a superlative verification strategy.

A logit mixed-effects model indeed confirms an interaction between SPEAKER-TYPE and COLOR RATIOS, such that COLOR RATIOS better predict the verification behavior of speakers who were classified as 'superlative' as opposed to those who were classified as 'proportional.' The model predicted percent-True from SPEAKER-TYPE and COLOR RATIOS for the first five data points on the left in the 3C MILDLY BALANCED condition (all ratios that are false under proportional truth conditions of *most*). The random effect structure was the maximal one supported by the design of Experiment 2 and included by-item intercepts and by-participant slopes and intercepts for SPEAKER-TYPE and COLOR RATIOS. The model also yielded a significant main effect of COLOR RATIOS, but this main effect was highly correlated with the interaction and we therefore do not put explanatory weight on it. All other correlations among fixed effects were within ±0.15. This is consistent with the results apparent in the graph in Fig. 10 and predicted by theories of *most* that treat it as a superlative construction. The fixed effects of the model are summarized in Table 5.

Finally, we note that if participants in the 'more than half' condition are classified using the same method that was used in (15) above for *most*, no superlative behavior is found in the data. This classification yields 94 subjects who are classified as "proportional" (*Mth-prop*) and 22 who are classified as "superlative" (*Mth-sup*). As

Table 5 Summary of the fixed effects for the 3C MILDLY BALANCED condition with *most*

Predictor	Coefficient	SE	z value	p
Intercept	−12.484	1.330	−9.385	**<0.001**
SPEAKER-TYPE	−2.846	2.088	−1.363	0.1728
COLOR RATIOS	9.731	1.133	8.591	**<0.001**
SPEAKER-TYPE × COLOR RATIOS	4.790	1.770	2.706	**<0.01**

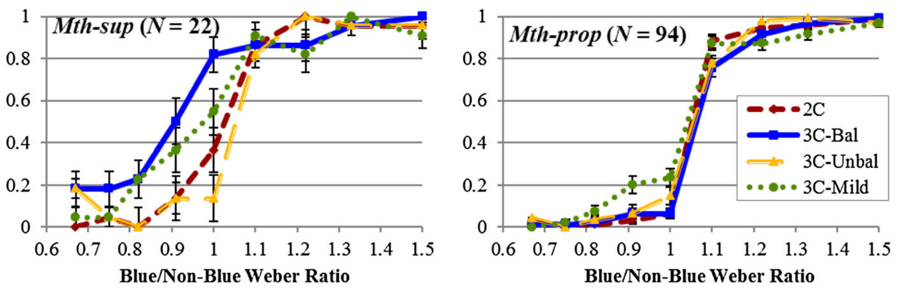

Fig. 11 Percent-True for "superlative" *more than half* (*Mth-sup*, *left*) and "proportional" *more than half* (*Mth-prop*, *right*) in Experiment 2. (Color figure online)

can be observed in Fig. 11, the "proportional" behavior in the graph on the right was made sharper by this classification. However, we find no indication of a superlative verification strategy for the participants who were classified as "superlative." Instead, we observe that the verification behavior for these speakers is generally more noisy; moreover, and importantly, we do not observe a selective increase in True responses to trials in the 'superlative' condition, namely all five trials in the 3C BALANCED condition, three trials in the 3C MILDLY BALANCED condition, and one trial in the 3C UNBALANCED condition.

A mixed-effects logit model predicting percent-True from SPEAKER-TYPE and COLOR RATIOS for the 3C MILDLY BALANCED condition for the *more than half* participants (parallel to the one described for the *most* participants above) confirms this finding. We find a main effect of COLOR RATIOS, but no effect of SPEAKER-TYPE and no interaction. That is, we observe only one type of speaker in this data, unlike the finding for *most*. The main effect of the COLOR RATIOS indicates that knowing the ratio improves the model predictions compared to having no information about the pictures at all. This is as expected, given that the COLOR RATIOS (correlated with superlative truth conditions) and the WEBER RATIOS (correlated with proportional truth conditions) converge for 6 out of the 9 ratios in the 3C MILDLY BALANCED condition. The fixed effects of this model are summarized in Table 6.

We can thus verify that the classification we have used above is parsimonious in grouping the noise in Experiment 2 together with the "superlative" data. Furthermore, as predicted by the truth conditions of *more than half*, no superlative behavior is observed even for those participants who were classified as "superlative" by the criterion in (15). Hence, the superlative behavior we observe for *most* is not

Table 6 Summary of the fixed effects for the 3C MILDLY BALANCED condition with *more than half*

Predictor	Coefficient	SE	z value	*p*
Intercept	−23.215	3.513	−6.609	**<0.001**
SPEAKER-TYPE	−6.263	5.434	−1.153	0.249
COLOR RATIOS	16.442	2.688	6.117	**<0.001**
SPEAKER-TYPE × COLOR RATIOS	5.803	3.906	1.486	0.137

an artifact of our classification of subjects or of the design of our experiment. Rather, it is genuinely attributable to the truth-conditional import of bare *most*. It is explained under a decompositional approach to *most*, whereas it is not predicted by the lexical ambiguity approach to *most*.

4 General discussion

In this section, we review the empirical results of Experiment 2 and their impact on the debate as to the correct analysis of *most*. We then discuss differences between *most* and *more than half* that emerge from both of our experiments, and, finally, we discuss the differences between the findings in our study and the findings of Lidz et al. (2011), who used a similar manipulation but did not find any evidence of sensitivity of *most* to the number of colors in the pictures against which the *most* statements were verified.

4.1 Proportional and superlative truth conditions for bare *most*

The goal of Experiment 2 was to test whether speakers can verify sentences like (16a) using superlative truth conditions as in (16c) in addition to the proportional truth conditions described in (16b).

(16) a. Most of the dots are blue
b. *Proportional truth conditions:*
$|\text{blue dots}| > |\text{non-blue dots}|$
c. *Superlative truth conditions:*
For each non-blue color Z, $|\text{blue dots}| > |\text{Z dots}|$

The data is quite clear. We saw that even when pooling data across all participants the rates of True responses to 'superlative' pictures (pictures for which the test statements were true under the reading described in (16c) but false under the reading in (16b)) show sensitivity to the COLOR manipulation, namely to the difference between the number of blue dots and the number of dots in the most numerous non-blue color (the blue-to-highest-non-blue ratio). This is a signature property of the superlative reading and thus indicates that at least some of our participants verified the pictures using the truth conditions as described in (16c). It cannot be explained if speakers had only access to a proportional reading, (16b), since under that reading there should only be sensitivity to the blue-to-non-blue ratio. Furthermore, we have seen that classifying participants into "superlative" and "proportional" speakers sharpens the data considerably: for the "superlative" group (about 40 % of our participants) the rates of True responses to superlative pictures mirror perfectly the white portion of Table 3, indicating that they have a strong preference to verify the arrays using a blue-to-highest-non-blue strategy. No such strategy was detected for the "proportional" group. We conclude from these observations that at least for our "superlative" speakers bare *most* in subject position is ambiguous between a proportional and a genuine superlative construal.

The fact that bare *most* in subject position can have a genuine superlative reading (in addition to the proportional reading) is expected under the view where *most* is in all its incarnations a superlative construction which gives rise to either proportional or superlative truth conditions depending on how the comparison class is set. It is not expected under the view where *most* is simply ambiguous between a lexical determiner with proportional truth conditions (bare *most*) and a superlative construction with superlative truth conditions (*the most*), since on that view bare *most* is always unambiguously proportional. Our data shows, however, that bare *most* is in fact ambiguous. To account for this, a lexical ambiguity approach would have to stipulate that bare *most* itself is ambiguous between a proportional meaning and a superlative meaning.[25] This proliferation of lexical ambiguity is an unwelcome consequence, especially since it offers no principled account of why *most* is ambiguous between these two meanings rather than any other possible (determiner) meanings or why *most* is ambiguous between a proportional and a superlative meaning while similar determiners such as *more than half* are unambiguously proportional.

In this connection, it is important to stress that the superlative-proportional ambiguity also arises in other languages, for example in German, as discussed in Hackl (2009), in Slavic, as discussed in Krasikova (2011), Szabolcsi (2012), Pancheva (forthcoming), and elsewhere. This fact, too, speaks against the lexical ambiguity account, as the latter offers no general perspective on the systematic (un)availability of proportional and superlative meanings for expressions that contain superlative morphosyntax. Moreover, just as in English, only determiners that contain a superlative morpheme seem to give rise to this ambiguity across languages: quantifiers such as *more than half* do not exhibit such an ambiguity.[26]

4.2 *Most* vs. *more than half*

In this section we return to the discussion of the asymmetric behavior of *most*, which is not apparent in the behavior of *more than half* in either Experiment 1 or

[25] For example, an anonymous reviewer suggests the following custom-tailored lexical ambiguity theory:

(i) $[\![\text{most}]\!](A)(B) = 1$ iff for all 'appropriate' $Y \subseteq A-B$, $|A \cap B| > |Y|$
– Dominant: Y can only be A−B itself
– Latent: Y can range over salient subsets of A−B

Note, however, that this proposal looks rather similar to the superlative-based treatment proposed in Hackl (2009) and summarized in (4b) and (5b) above. It differs only in that no reference to degrees or to the degree quantifier expressed by *-est* is acknowledged. It thus prompts the question why its meaning is related to the meaning of the superlative construction *the most.* Furthermore, this account does not give a principled explanation as to why only these two interpretations are attested—in particular, why interpretations such as Y = the smallest subset of A–B, or Y = two or more subsets of A–B, are unattested, despite being in principle sensible.

[26] To be sure, the decompositional view of *most* faces remaining challenges, as noted by an anonymous reviewer. In particular: (i) the presence/absence of the definite article; (ii) the tendency of *most* but not *the most* to have a generic interpretation when combining with a plural noun, and to prefer a partitive construction in episodic sentences. These issues are beyond the scope of this paper, but see Pancheva (to appear) for a recent attempt to explain the interaction with the definite article.

Experiment 2. In both experiments we saw that in the 2C condition,[27] the verification behavior of *more than half* resembles a step function that changes sharply from True rates close to zero for all 'false' trials to True rates close to 100 % for all 'true' trials. For *most*, we observe parallel low True rates for all 'false' trials, but the True rates are lower than those of *more than half* for all 'true' trials, Furthermore, the first 'true' ratio for *most* yields True judgments only 77 % of the time, while other ratios yield True judgments 90 % of the time or more. Although never formally recognized or explained in any previous literature on *most*, this asymmetry has been observed for *most* in prior experimental work (Yosef Grodzinsky, p.c.).

We believe that this finding can be related to an observation regarding the pragmatics of the use of *most* and *more than half*. In particular, *more than half* is used for proportions that are above but still close to 50 %, while *most* tends to imply proportions that are significantly higher than 50 % (Peterson 1979; Westerstahl 1985; Horn 2005; cf. Ariel 2004 for a similar observation for the Hebrew *rov* 'most').[28] This observation has recently been verified through corpus work by Solt (to appear), who examined the use of *most* and *more than half* in the Corpus of Contemporary American English (COCA; Davies 2008) and found that *more than half* was typically used to describe percentages between 50 and 65 %, while *most* was rarely used for percentages below 60 %. Below are some examples from the COCA corpus, taken from Solt (to appear).

(17) a. The survey showed that *most* students (81.5 %) do not use websites for math-related assignments. (*Education* 129(1), 56–79, 2008)
b. *Most* respondents—63 %—said the best movie for date night is a comedy. (*Redbook* 208(6), 158, 2007)
c. *Most* Caucasian grandparents were married (67 %), had attained an education level above high school (64 %), and lived on an annual household income above $20,000. (74 %). (*Journal of Instructional Psychology* 24(2), 119, 1997)

(18) a. *More than half* of respondents (55 %) say that making money is more important now than it was five years ago. (*Money* 21(3), 72, 1992)
b. *More than half* of the respondents (60 %) earned Ph.D. degrees. (*Physical Educator* 53(4), 170, 1996)
c. And while *more than half* of us grill year-round (57 %), summertime is overwhelmingly charcoal time. (*Denver Post*, May 24, 2000)

Solt (to appear) describes several additional differences between *most* and *more than half* that stem from the kinds of nouns that these two determiners normally combine with (kind vs. group denoting), the overall higher frequency in the corpus

[27] The 2C condition is the only condition where there is no interference from the Color manipulation. Because of this complicating interference in the other Color conditions, we restrict the discussion above to the 2C condition. However, the same observation holds for all conditions for participants who we classified as "proportional" *most* speakers in Sect. 3.3.

[28] Here we will not discuss previous attempts to build this property of *most* into its semantics.

of *most* compared to *more than half*, and their normal usage to describe generic vs. 'survey result' readings, respectively. Solt also shows that *most* is used more often with vague and uncountable domains, while *more than half* appears less compatible with such domains, as illustrated in (19) (also from Solt, to appear).[29]

(19) a. But like *most* things, obesity is not spread equally across social classes.
b. ?But like *more than half* of things, obesity is not spread equally across social classes.

Solt proposes a formal account of these differences between *most* and *more than half* based on the interaction of their logical forms with different measurement scale structures.

We see, then, that although *most* and *more than half* are truth-conditionally equivalent, they are used in different contexts. This finding is consistent with the results of Experiments 1 and 2. First, both determiners are false under the exact same conditions, and hence we expect them to be judged as false at very high rates for all 'false' conditions. However, while the two determiners are true in all the same cases, they are not used in the same way: *more than half* is more often used for proportions that are close to 50 %, and *most* is more often used for proportions significantly higher than 50 %. This fact may have contributed to the difference in rates of True judgments for *most* and *more than half*: the rate of True responses to our trials may reflect the felicity of the corresponding statement in the context.

As we have seen, participants verify *more than half* statements as true at very high rates for all 'true' ratios, including those that are very close to a Weber ratio of 1: these are contexts in which *more than half* is often used felicitously. On the other hand, *most* is used less often in such contexts, and hence *most* is judged as true less often in these cases. As the Weber ratio increases, the percentage of speakers who find the *most* statement felicitous increases too. Furthermore, since the highest proportion of blue dots found in our pictures was only 60 % of the total number of dots, we can understand why even at the highest Weber ratio in our experiments, *more than half* is consistently judged as true more often than *most* for all Color conditions: as Solt shows, 60 % is still within the range of percentages for which speakers tend to prefer to use *more than half* as opposed to *most*.

To conclude, the asymmetry property of *most* seems to be related to the pragmatics of its usage. Although *most* is true in the same cases as *more than half* (under proportional truth conditions), it is less often used to describe pictures with the proportions that were found in our experiments, compared to *more than half*. Furthermore, it is predicted that both determiners will be clearly judged as false in all 'false' pictures because the conditions of their use entail that both *most* and *more than half* statements are infelicitous in such cases. Finally, we note that our data does not speak to the question of whether the distributional facts of *most* are the underlying cause of the asymmetry we observed or whether the semantic properties

[29] For related observations about *most* and *the most* and their counterparts in Hungarian, Russian, and German see Szabolcsi (2012).

of *most*, e.g. along the lines of Solt (to appear), are responsible for both the distributional data and the asymmetry in our results.[30]

4.3 Comparison with Lidz et al. (2011)

In this section we review the results of a previous sentence verification study of *most* conducted by Lidz et al. (2011), which, even though similar to our study in many ways, obtained different findings than those we reported above.[31] Although it was not the explicit goal of Lidz et al. to test whether bare *most* has a superlative reading in subject position, their study directly bears on the question of whether such a reading exists. In particular, while we have found that *most* is ambiguous between a dominant proportional reading and a latent superlative reading, the results of Lidz et al. suggest that *most* is unambiguous and has only the proportional reading. Below we survey Lidz et al.'s study and investigate potential sources for the conflict between their findings and ours. We argue that the source of the difference lies in the different task demands and in the way the results of the two studies were interpreted.

4.3.1 A brief summary of Lidz et al. (2011)

Lidz et al. (2011) examine the verification of the question *Are most of the dots blue?* with respect to dot arrays of varying complexities. Arrays varied along two dimensions: (a) the Weber ratio of blue:non-blue dots (1:2, 2:3, 3:4, and 4:5; in half of the cases the blue dots outnumbered the non-blue dots); and (b) the number of colors used in the arrays (2, 3, 4, or 5). The goal of the study was to test a prediction of the standard Generalized Quantifier Theory (GQT) approach to *most*, given in (20), when combined with the hypothesis that when speakers have to determine the truth/falsity of a sentence, they are biased towards verification strategies that employ operations specified as part of the truth-conditional import of the statement—Lidz et al.'s Interface Transparency Thesis (ITT), given in (21).

(20) $[\![\text{most}]\!](A)(B) = 1$ iff $|A \cap B| > |A - B|$ Standard GQT

(21) *Interface Transparency Thesis* (*ITT*; Lidz et al. 2011, p. 234):
The verification procedures employed in understanding a declarative sentence are biased towards algorithms that directly compute the relations and operations expressed by the semantic representation of that sentence.

[30] An anonymous reviewer asks whether *the most* exhibits the same asymmetry that bare *most* does. If only one determiner exhibits this property but not the other, the decompositional approach to *most* will face a challenge explaining this fact. We note that Kotek et al. (2011b) show that outside of a particular syntactic configuration, *the most* in subject position is infelicitous for many English speakers, making it impossible to test *the most* in the same experimental setting used for bare *most* in the present paper.

[31] The results of Lidz et al. (2011) also differ from the findings in Kotek et al. (2011a), which were consistent with the findings of our Experiment 2: that bare *most* in subject position is ambiguous between a dominant proportional reading and a latent superlative reading.

Lidz et al. (2011) argue that given (20) and the ITT, speakers are expected to employ a verification strategy that enumerates the blue dots and the non-blue dots and compares the two numbers. For this to be feasible, they argue, it is necessary to select and attend to the non-blue set. However, it has been shown that (at least under extremely short exposure times) it is impossible for the visual system to select a set of objects based on a "negative" feature such as non-blue (Wolfe 1998; Treisman and Gormican 1988; Treisman and Souther 1985; see discussion in Lidz et al. 2011). Instead, in order to select the non-blue set, it is necessary to select the subsets of homogeneously colored dots that make up the non-blue set. Lidz et al. reason that this step of selecting each subset of homogeneously colored dots should be sensitive to the number of colors used in the arrays. Therefore, if determining the cardinalities of all the homogeneously colored subsets of non-blue dots is part of the verification procedure of *most* statements, increasing the number of colors in the arrays should make the verification of *most* increasingly difficult.

To test this prediction, Lidz et al.'s study presented their dot arrays for only 150 ms. This made a verification strategy based on counting all the dots individually impossible; it also made it impossible for the visual system to directly select the non-blue set. Instead, participants had to rely on estimating the cardinality of whatever sets they used in the verification process. Importantly, estimating the cardinality of a set is a process whose accuracy is governed by the Weber ratio and for which there exist well-established psychophysical models (e.g., Pica et al. 2004). Moreover, as Halberda et al. (2006) have shown, the number of sets whose cardinality participants can successfully estimate at such brief exposure times is limited: participants can estimate the cardinality of the set of all dots and, in addition, of up to two homogenously colored subsets of dots. Thus, when a task requires estimating more than two subsets in addition to the total set, performance drops off markedly. Based on this observation, Lidz et al. reasoned that verifying *most* statements using a strategy that requires estimating the size of each homogenously colored subset of dots individually should be markedly more difficult; it should no longer be predicted by the Weber ratio when the number of colors used in the array is 3, 4, or 5.[32] However, this is not what they found in their study. Rather, their results, shown below in Fig. 12 from Lidz et al. (2011), indicate a marked lack of sensitivity to the number of colors in the array, as well as a strong dependency on the Weber ratio.

Since their results show no effect of Color, Lidz et al. (2011) conclude that the composition of the non-blue set could not have played a direct role in the verification of the *most* statements. To account for this insensitivity to the number of colors used in the arrays, Lidz et al. propose the analysis of *most* in (22b) in place of the more traditional GQT treatment in (22a).[33]

[32] Whether performance should break down when there are 3 colors in the array or only when there are 4 or more is somewhat unclear given the results in Halberda et al. (2006). Whatever the answer to this question may be, it does not change the interpretation of the experiment in Lidz et al., since there was no effect of color across all levels.

[33] To make this argument, Lidz et al. rely on additional assumptions that provide a bridge between the two hypotheses in (22) and how speakers verify statements in general. An in-depth discussion of these assumptions would lead us too far afield here, and it is peripheral to our main argument—that *most* must be analyzed as a superlative structure—because both (22a) and (22b) are inconsistent with the superlative behavior observed in our study.

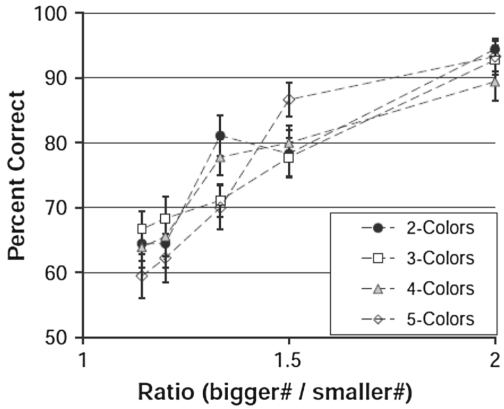

Fig. 12 Results of the *most* experiment in Lidz et al. (2011). Reproduced with the authors' permission

(22) a. $[\![\text{most}]\!](A)(B) = 1$ iff $|A \cap B| > |A - B|$ Standard GQT
b. $[\![\text{most}]\!](A)(B) = 1$ iff $|A \cap B| > |A| - |A \cap B|$ Lidz et al. (2011)

Lidz et al. (2011) argue that under the ITT, (22b) differs from (22a) in that it only requires determining the cardinality of all the dots in the array (|A|) and the cardinality of the blue set (|A ∩ B|). With these two quantities established, the cardinality of the non-blue set can now be calculated as |A| − |A ∩ B|. That is, the non-blue set need not be estimated as the sum of the cardinalities of all the homogeneously colored non-blue subsets.

The specific assumptions underlying this argument deserve closer scrutiny; we will return to them in Sect. 4.3.3. For now, we simply point out that the lack of an effect of Color in Lidz et al.'s (2011) experiment implies that bare *most* could not have been verified using a superlative verification strategy. Such a strategy would have required the estimation of each homogeneously colored subset of the non-blue set separately, and that behavior is predicted to be sensitive to Color. Instead, only proportional truth conditions could have been available to participants in Lidz et al.'s experiment. This appears to be in conflict with the present study. As our experiment shows, superlative truth conditions are in principle available to some speakers who verify *most* statements. Lidz et al.'s proposal (just like the standard GQT treatment), however, only allows for proportional truth conditions. As such, it is a version of the lexical view of *most* and so cannot straightforwardly account for the superlative behavior found in our results.

4.3.2 The role of task demands in sentence verification

In this section we discuss key differences between the study reported here and Lidz et al.'s (2011) work and argue that the specific design and task demands of the experiment in Lidz et al. played a critical role in obtaining the result that the color manipulation had no effect. This, in turn, casts doubt on the generalizability of the results and on the robustness of the conclusions drawn from that study.

Recall that both studies explored the verification of bare *most* in subject position through experiments that manipulated the COLOR and WEBER RATIOS of the dots in the pictures against which a *most* statement was verified. The present study also used a DETERMINER manipulation to compare the behavior of *most* to that of *more than half*. Lidz et al. (2011) found an effect of WEBER RATIOS but no effect of COLOR, insofar as all four COLOR conditions in their experiment were verified with similar levels of accuracy. The present study found an effect of WEBER RATIOS as well, but in addition it found an effect of COLOR for *most* but not *more than half*, such that *most* statements were consistently judged as true significantly more often for 'superlative' pictures than for pictures with the same Weber ratios that were false under the superlative reading. A similar effect of COLOR was also found in Kotek et al. (2011a), who used a COLOR manipulation with Weber ratios very close to 1 and found that pictures that contained three colors and were true only under the superlative reading were judged as true more often than corresponding pictures with the same ratios that contained just two colors. This result led Kotek et al. (2011a) as well as the present study to adopt a view of *most* that is compatible with the structural view of *most*, under which *most* is the superlative form of *many* (Hackl 2009).

We see, then, that both Lidz et al. and the present study use a COLOR manipulation and a WEBER RATIOS manipulation. However, only the present study, following Kotek et al., also uses a DETERMINER manipulation; this manipulation helps to ensure that whatever results are obtained for *most* can be attributed to the semantics and verification procedures associated with *most*, rather than some general difficulty related to the design of the experiment or to its participants. If experimental design were the issue, we would expect the verification of *more than half* to be affected in a similar manner to that of *most*. Similarly, if *most* is unambiguously a proportional determiner and truth-conditionally equivalent to *more than half*, as predicted by the lexical view of *most*, we would expect any experimental manipulation to affect these two determiners equally.

One important source of difference between the experiments is the mode of presentation of dot arrays in them. We saw that the experiments in the present study allowed participants unlimited time to make their decision. The Lidz et al. experiment, on the other hand, used a flash presentation method: participants saw the dot arrays for 150 ms and were asked to answer the question *Are most of the dots blue*? based on whatever information they could gather within that time frame. Moreover, participants were always asked the very same question: *Are most of the dots blue*?, where the color of the dots was not varied and no fillers were used, for 400 trials in all. We would like to suggest that this design and the specific task demands it imposed on participants biased them toward using a verification strategy such as the one derivable from combining the ITT with (22b). The flash presentation allowed participants to gather enough information to support a verification strategy compatible with proportional truth conditions in all of the experimental conditions, irrespective of whether there were 2, 3, 4, or 5 colors in a given array. A superlative verification strategy requiring estimates for the number of dots in each color, by contrast, was only supported in half of the conditions—when the arrays contained 2 or 3 colors but not when they contained 4 or 5 colors.[34] Given that participants encountered the very same question 400 times in a row, it is likely that they acted strategically and adopted a verification procedure that could guarantee a successful verification of the question in all cases—that is, a strategy compatible with proportional truth conditions.[35] Thus, we think the extreme task demands imposed on participants in this study cast doubt on whether the verification strategy that was observed indeed reflects the underlying components of the truth conditions of *most* and the way that they are stated, as assumed by Lidz et al. Instead, the behavior displayed by participants might reflect a less transparent translation of the truth conditions of *most* into a strategy that could guarantee successful verification of the *most* statement in all experimental conditions, regardless of the specific composition of a given dot array. Thus, while the proposal in (22b) might indeed correctly reflect the behavior of participants in Lidz et al.'s study, it is not obvious that it should be viewed as reflecting the underlying truth conditions of *most* outside of Lidz et al.'s experimental settings.

To put it differently, Lidz et al.'s argument that the semantic import of *most* is to be stated as in (22b) rather than as in (22a) or, for that matter, in terms of a superlative semantics as we argue here, relies on stating the ITT in a way that does not take into account the specific task demands under which speakers might engage in verification. This is unrealistic. Moreover, it can create the false impression, as is the case here, of conflicting data. A more realistic version of the ITT, which avoids drawing inconsistent inferences from experimental data about the underlying semantics of an expression, would make the ITT sensitive to the task demands. Thus, minimally an amendment of the following sort is in order:

(23) *The modified Interface Transparency Thesis* (*mITT*)

When determining the truth/falsity of a statement in a given situation, speakers exhibit a bias towards using verification procedures that employ operations specified as part of the truth-conditional import of the statement, as supported by the task demands brought about by that situation.

An amendment of the ITT to include sensitivity to specific task demands makes inferring semantic properties of expressions from verification data less direct than Lidz et al. assumed. In particular, for a given linguistic expression there may be several kinds of "default" verification strategies and speakers might choose among those according to the specific experimental conditions. Thus, for any given experimental setting, verification data can be explained only by combining

[34] Recall that following work by Halberda et al. (2006), in conditions like those that obtained in the experiment of Lidz et al. (2011) participants are able to estimate the total number of dots in the array, and additionally the sizes of up to two homogeneously colored subsets of dots.

[35] An anonymous reviewer proposes further that given the lack of fillers in the Lidz et al. (2011) study, it is possible that participants in that study interpreted the task as simply deciding whether or not there were more blue dots than other dots in the arrays presented to them, effectively ignoring the prompt they were given.

hypotheses about the semantics of the studied expression with an (ideally independently justified) theory of the task.[36]

Conflicting experimental results represent a challenge, but we argue that in the present case a resolution can be found by assuming that the underlying semantics of *most* is superlative in nature. Under this assumption, the results of the present study (as well as those reported in Kotek et al. 2011a) are straightforwardly explained, while the results of Lidz et al. (2011) can be accounted for under the additional assumption that speakers strategically worked with the proportional reading, which, following Lidz et al., can be represented as in (22b). By contrast, on the view that the semantics of *most* is underlyingly as in (22b), only the data in Lidz et al. is explained. The superlative behavior we found in the present study is not understandable on those terms alone since the task demands of our study do not impinge on a verification strategy that is based on (22b). Thus, there is, prima facie, no reason to expect that speakers should not employ the same verification strategy that they did in Lidz et al.'s experiment.

4.3.3 The asymmetry of most and ANS models

In the previous subsection we argued that the main inconsistency between the present study and the Lidz et al. study—the presence/absence of an effect of Color—is resolved once we assume a superlative semantics for *most* in conjunction with a modified ITT, which allows for the specific task demands in a verification setting to influence which verification procedure is chosen by speakers as the strategically most advantageous. The current section briefly discusses another, somewhat less critical inconsistency between the two studies.

Recall from the discussion of Experiment 1 that in the case of *most* we observe an asymmetry between the rates of True responses for Weber ratios above 1 and the rates of False responses for Weber ratios below 1: while the latter are close to zero for all relevant ratios (and are no different from the rates of no responses for *more than half* for all those Weber ratios) the former are lower than the corresponding True rates for *more than half*, for all ratios except the highest one. Moreover, we saw that the rate of True responses for Weber ratio 1 was close to zero, indicating that speakers were confidently judging the *most* statement to be false in this case.

Neither observation is apparent in Lidz et al.'s reported data (see Fig. 12), since they assumed that the verification process of *most* statements can be faithfully modeled as a symmetric discrimination task between two quantities represented by the Approximate Number System (ANS).[37,38] Under this model, it is possible to assume that the rate of True responses at Weber ratio 1 is 50 %, since two equal

[36] This is, of course, generally true for all behavior that linguistic theories aim to explain, and not just for sentence verification paradigms (cf. Marantz 2005).

[37] The ANS is an evolutionarily ancient cognitive mechanism that is able to generates an approximate representation of the number of items in a set, in accord with Weber's law: the discriminability of two quantities is a function of their ratio; discriminability increases as the Weber ratio increases.

[38] Modeling the verification task as a symmetric discrimination task means, informally speaking, that participants were assumed to determine in each verification trial whether $n>m$ or $m>n$ (with n representing the estimated cardinality of blue dots and m the estimated cardinality of the non-blue dots).

ANS quantities are maximally confusable. Moreover, it is now legitimate to collapse response rates for Weber ratios and their multiplicative inverses, as Lidz et al. indeed do, to produce the data that was subsequently modeled using an independently provided psycho-physical model of the ANS. However, the fact that our results do not support these assumptions indicates that the use of ANS modeling as the basis for the study of *most* may not be justified in the general case.

5 Conclusion

This paper has introduced new evidence regarding the correct semantics of bare *most*. Previous sentence verification studies of *most* have yielded diverging results: Kotek et al. (2011a) find that *most* is ambiguous between a dominant proportional reading and a latent superlative reading, while the findings in Lidz et al. (2011) are consistent with *most* being unambiguously proportional and truth-conditionally equivalent to *more than half*.

Here, we presented the results of a study that combined the manipulations of the two previous studies and showed that bare *most* does indeed have a superlative reading in subject position. In particular, we find an effect of the number of colors in the pictures for *most* but not for *more than half* precisely in those pictures where the test sentences were true under superlative truth condition (but false under proportional truth conditions). This reflects the fact that *most*, but not *more than half*, can be verified using a verification strategy that is compatible with superlative truth conditions.

This result is compatible with a decompositional analysis of *most*, according to which both *the most* and bare *most* are built from the same basic ingredients—a gradable predicate *MANY* and the superlative operator *-est*—but project different LFs, as laid out e.g. in Hackl (2009). The results are unexpected under the lexical view of *most*, where bare *most* and *the most* are unrelated lexical items, with bare *most* exclusively expressing proportional semantics and *the most* exclusively expressing superlative semantics.

We explained the differences between the findings of our study and those of Lidz et al. (2011) study as stemming from differences in the task demands, which may have biased participants in the latter study towards using a verification strategy that is only compatible with proportional truth conditions. We have argued that such a strategy is guaranteed to succeed in all trials in Lidz et al.'s study, while a strategy compatible with superlative truth conditions is expected to fail in 50 % of the trials. This makes it unsurprising that no evidence of a superlative reading of bare *most* was found in that earlier study. But it follows, then, that the absence of any evidence indicative of a superlative construal of bare *most* cannot be taken to mean that bare *most* has only a proportional meaning.

Acknowledgments We would like to thank Adam Albright, Edward Flemming, Danny Fox, the participants of the Syntax and Semantics Experimental Lab at MIT, and two anonymous *NALS* reviewers for valuable comments. We would also like to thank Jeffrey Lidz and his coauthors for their willingness to answer questions about their 2011 study and for allowing us to reproduce their figure.

References

Ariel, Mira. 2004. Most. *Language* 80: 658–706.

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.

Barr, Dale, Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3): 255–278.

Barwise, Jon, and Robin Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 4: 159–219.

Bates, D.M., and M. Maechler. 2009. lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-32.

Bresnan, Joan. 1973. Syntax of the comparative clause construction in English. *Linguistic Inquiry* 4(3): 275–344.

Davies, Mark. 2008. The corpus of contemporary American English (COCA): 400+ million words, 1990–present. Online at http://americancorpus.org.

Erlewine, Michael Yoshitaka, and Hadas Kotek. to appear. A streamlined approach to online linguistic surveys. *Natural Language and Linguistic Theory* (to appear).

Hackl, Martin. 2009. On the grammar and processing of proportional quantifiers: *most* versus *more than half. Natural Language Semantics* 17: 63–98.

Halberda, Justin, Sean F. Sires, and Lisa Feigenson. 2006. Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science* 17: 572–576.

Heim, Irene. 1985. Notes on comparatives and related matters. Unpublished manuscript, University of Texas at Austin. Semantics Archive. http://semanticsarchive.net/Archive/zc0ZjY0M.

Heim, Stefan, Katrin Amunts, Dan Drai, Simon B. Eickhoff, Sarah Hautvast, and Yosef Grodzinsky. 2012. The language–number interface in the brain: a complex parametric study of quantifiers and quantities. *Frontiers in Evolutionary Neurosciences* 4: 1–12.

Horn, Laurence R. 2005. The border wars: A neo-Gricean perspective. In *Where semantics meets pragmatics*, ed. Klaus von Heusinger and Ken Turner, 21–48. Amsterdam: Elsevier.

Kotek, Hadas, Yasutada Sudo, Edwin Howard, and Martin Hackl. 2011a. Most meanings are superlative. In *Syntax and Semantics 37: Experiments at the interfaces*, ed. Jeff Runner, 101–145. Bingley: Emerald Group Publishing.

Kotek, Hadas, Yasutada Sudo, Edwin Howard, and Martin Hackl. 2011b. Three readings of most. In *Proceedings of SALT 21*, ed. Neil Ashton et al., 353–372. Ithaca: CLC Publications.

Krasikova, Sveta. 2011. Definiteness in superlatives. In *18th Amsterdam Colloquium Pre-proceedings*, 404–414. Amsterdam: ILLC. http://www.illc.uva.nl/AC/AC2011/Proceedings/.

Lidz, Jeff, Paul Pietroski, Tim Hunter, and Justin Halberda. 2011. Interface transparency and psychosemantics of *most*. *Natural Language Semantics* 19: 227–256.

Link, Godehard. 1983. The logical analysis of plurals and mass terms: A lattice-theoretical approach. In *Meaning, use, and interpretation of language*, ed. Rainer Bäuerle et al., 302–323. Berlin: de Gruyter.

Marantz, Alec. 2005. Generative linguistics within the cognitive neuroscience of language. *The Linguistic Review* 22: 429–445.

Matthewson, Lisa. 2001. Quantification and the nature of cross-linguistic variation. *Natural Language Semantics* 9: 145–189.

Pancheva, Roumyana. To appear. Quantity superlatives: The view from Slavic and its cross-linguistic implications. In *Proceedings of CLS 49*. Chicago: The Chicago Linguistic Society.

Peterson, Philip L. 1979. On the logic of "few", "many", and "most". *Notre Dame Journal of Formal Logic* 20: 155–179.

Pica, Pierre, Cathy Lemer, Veronique Izard, and Stanislas Dehaene. 2004. Exact and approximate arithmetic in an Amazonian indigene group. *Science* 306: 499–503.

Pietroski, Paul, Jeff Lidz, Tim Hunter, and Justin Halberda. 2009. The meaning of *most*: Semantics, numerosity and psychology. *Mind & Language* 24(5): 554–585.

Pietroski, Paul, Jeff Lidz, Justin Halberda, Tim Hunter, and Darko Odic. 2011. Seeing what you mean, mostly. In *Syntax and semantics 37: Experiments at the interfaces*, ed. Jeff Runner, 187–224. New York: Academic Press.

Pinkham, Jessie. 1985. *The formation of comparative clauses in French and English*. New York: Garland Publishing.

Solt, Stephanie. 2011. How many *mosts*. In *Proceedings of Sinn und Bedeutung 15*, ed. Ingo Reich et al., 565–579. Saarbrücken: Saarland University Press.
Solt, Stephanie. To appear. On measurement and quantification: The case of *most* and *more than half*. *Language*.
Szabolcsi, Anna. 1986. Comparative Superlatives. In *Papers in Theoretical Linguistics (MITWPL 8)*, ed. Naoki Fukui et al., 245–265. Cambridge, MA: MIT.
Szabolcsi, Anna. 2010. *Quantification*. Cambridge: Cambridge University Press.
Szabolcsi, Anna. 2012. Compositionality without word boundaries: (The) more and (the) most. In *Proceedings of SALT 22*, ed. Anca Chereches, 1–25. Ithaca: CLC Publications.
Tomaszewicz, Barbara M. 2011. Verification strategies for two majority quantifiers in Polish. In *Proceedings of Sinn und Bedeutung 15*, ed. Ingo Reich et al. Saarbrücken: Saarland Unversity Press.
Treisman, A., and S. Gormican. 1988. Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review* 95(1): 15–48.
Treisman, A., and J. Souther. 1985. Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: Human Perception & Performance* 16(3): 459–478.
Westerstahl, Dag. 1985. Logical constants in quantifier languages. *Linguistics and Philosophy* 8: 387–413.
Wolfe, J.M. 1998. Visual search. In *Attention*, ed. H. Pashler, 13–73. London: University College London Press.
Yabushita, Katsuhiko. 1999. The unified semantics of *mosts*. In *Proceedings of WCCFL 18*, ed. Sonya Bird et al., 320–334. Somerville: Cascadilla Press.