Beyond Truth Conditions:
an investigation into the semantics of 'most'

Paul Pietroski*, University of Maryland*

Justin Halberda*, Johns Hopkins University*

Jeffrey Lidz, *University of Maryland*

Tim Hunter*, University of Maryland*

Contact Info:
Paul Pietroski
Department of Linguistics
University of Maryland
Marie Mount Hall
College Park, MD 20742 USA
Email: pietro@umd.edu
Phone: +1 301-395-1747

Word Count Text:

Word Count Method and Figure Legends:

Title Character Count:

Keywords: analog magnitude, approximate number system, sets, number, quantifiers, Hume, mathematics, most, some, every, all, more, language development

Abstract

To a first approximation, 'Most of the dots are yellow' means that the number of yellow dots exceeds the number of nonyellow dots. But given current semantic theories, there are many ways to describe the meaning of 'most'. We offer a framework for distinguishing these hypotheses, and we present data that tells against the initially compelling idea that 'most'—along with many expressions of cardinality—is understood in terms of one-to-one correspondence.

English-speaking adults were asked to evaluate 'Most of the dots are yellow', as true or false, on many trials in which both yellow dots and blue dots were shown briefly on a computer screen. Displays were manipulated to vary the ease of applying a "one-to-one with remainder" strategy, in evaluating the sentence, and the ease of using the Approximate Number System as part of a strategy involving comparisons of (approximations of) cardinalities. Results show no evidence of participants evaluating the target sentence in terms of one-to-one correspondence.

The quantificational word 'most' is discussed as a case study for connecting proposals in formal semantics, usually motivated by intuitions about sentential truth conditions, with experimental methods that can reveal the representations deployed when speakers understand words.

How is the word ‘most’ related to human capacities for detecting and comparing numerosities? One might think the answer is both obvious and explicit in standard semantic theories: ‘most’ is understood in terms of a capacity to compare cardinal numbers; for example, ‘Most of the dots are yellow’ means that the number of yellow dots is greater than the number of nonyellow dots. But this is not at all obvious. There are many possibilities to consider, with regard to how competent speakers might understand ‘most’, and we discuss experimental evidence that tells against some initially attractive hypotheses. Our goal, however, is not merely to defend a certain view about one lexical item. We want to illustrate how semantics and psychology can and should be pursued in tandem, especially with regard to the capacities that let humans become numerate.

Following common practice in semantics, we characterize lexical meanings formally, often in set-theoretic terms. Though as reviewed in section one, there are many ways—indeed, many truth conditionally equivalent ways—to describe the meaning of ‘most’. Our aim is to give some of these distinctions empirical bite, in a way that permits adjudication among alternative psychological hypotheses. Of necessity, this discussion covers more than one field, with consequent expository perils. Readers from different disciplines will have equal opportunities to be at home and at sea, as the sections below deal with formal semantics, experimental psychology, and attempts to forge some connections. The goal is to link questions about the semantic contribution of ‘most’, in complex expressions, to questions about how that contribution is naturally represented in speakers who somehow compute the meanings of linguistic expressions.

1. Background Semantics

In this section, we survey some textbook analyses of ‘most’. The focus on truth conditions, standard among semanticists, may be less familiar to some readers. But in addition to providing necessary background for what follows, the review lets us stress the availability of formally distinct meaning specifications that are truth conditionally equivalent. And this highlights our main questions, concerning the mental representations that speakers associate with ‘most’. At the end of this section and thereafter, we urge an explicitly cognitive conception of semantics, according to which the goal of recursively specifying truth conditions for sentences is part of a larger attempt to describe the representations and procedure(s) that speakers implement in understanding expressions of natural language.

In the scene shown in Figure 1, most of the dots are yellow.

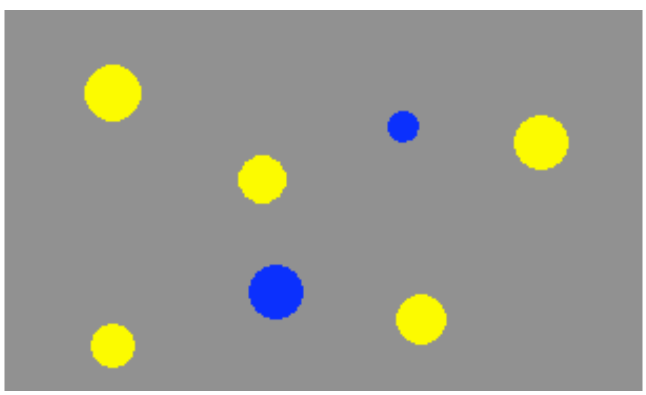


Figure 1

It is also the case, in this scene, that the number of yellow dots is greater than the number of dots that are not yellow. This correlation is not accidental. However many dots there are, necessarily, most of the dots are yellow if and only if (iff) the yellow dots outnumber the other dots. Moreover, that necessity seems obvious. So one might propose that (1)

(1) Most of the dots are yellow

means just this: $\#\{x: \text{Dot}(x) \ \& \ \text{Yellow}(x)\} > \#\{x: \text{Dot}(x) \ \& \ \neg\text{Yellow}(x)\}$; where

‘#{ ... }’ indicates the cardinality of the set in question. In our scene, $\#\{x: Dot(x) \,\&\, Yellow(x)\} = 5$, and $\#\{x: Dot(x) \,\&\, \neg Yellow(x)\} = 2$. So given that $5 > 2$, the condition imposed by the hypothesized meaning of sentence (1) is satisfied. More generally, one might say that a sentence of the form ‘Most (of the) Δs are Ψ’ means that the number of Δs that are Ψ is greater than the number of Δs that are not Ψ.[i]

On this view, the determiner ‘most’ signifies a relation that one set can bear to another. This allows for a unified semantics of determiners—words like ‘every’ and ‘some’, which can combine with a noun and then a tensed predicate to form a complete sentence as in (2-4).[ii]

(2) [(Every dot) (is yellow)]

(3) [(Some dot) (is yellow)]

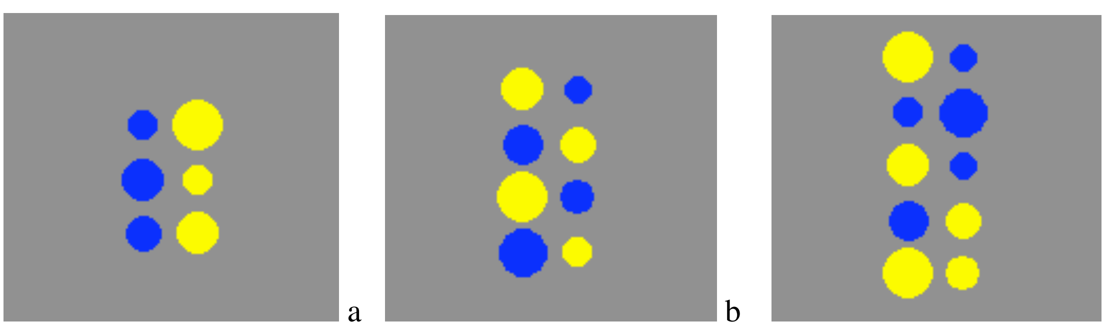
(4) [(Five dots) (are yellow)]

One can say that ‘every’ signifies the subset relation, and hence, that (2) is true iff $\{x: Dot(x)\} \subseteq \{x: Yellow(x)\}$. If ‘some’ signifies intersection, (3) is true iff $\{x: Dot(x)\} \square \{x: Yellow(x)\}$. But one can also describe the semantic contributions of ‘every’ and ‘some’ in terms of cardinalities: every Δ is Ψ iff $\#\{x: \Delta(x) \,\&\, \neg\Psi(x)\} = 0$; some Δ is Ψ iff $\#\{x: \Delta(x) \,\&\, \Psi(x)\} > 0$. From this perspective, ‘five’ is a special case of ‘some’, since (exactly) five Δs are Ψ iff $\#\{x: \Delta(x) \,\&\, \Psi(x)\} = 5$. This invites the characterization of ‘most’ suggested above, according to which most Δs are Ψ iff $\#\{x: \Delta(x) \,\&\, \Psi(x)\} > \#\{x: \Delta(x) \,\&\, \neg\Psi(x)\}$. But while these biconditionals may correctly and usefully describe the truth conditional contributions of determiners, questions about meaning and understanding remain.

Prima facie, a speaker can understand ‘every’ and ‘some’ without having a concept of zero, or any capacity to determine the cardinality of a set. Put another way, it seems like overintellectualization to say that sentence (2) *means* that zero is the number of dots that are not yellow, or that sentence (3) *means* that the set of yellow dots has a cardinality greater than zero.[iii] Indeed, one might think that the semantic contributions of ‘every’ and ‘some’ should be characterized in terms of simple first-order logic—every Δ is Ψ iff $\forall x{:}\Delta x(\Psi x)$, some Δ is Ψ iff $\exists x{:}\Delta x(\Psi x)$—and not in terms of relations between sets at all. We will not pursue this particular issue here. But analogous issues arise with regard to ‘most’.

One *can* capture the truth conditional contribution of ‘most’ without reference to numbers, at least for examples involving finitely many things. Instead of appealing to cardinalities, one can appeal to one-to-one correspondence. And this raises questions about how theorists *should* describe the truth conditional contribution of ‘most’ if their aim is to characterize the mental representations that determine how speakers understand complex expressions involving this word.

The notion of cardinality is intimately related to the notion of one-to-one correspondence. Upon reflection, this is obvious. The yellow dots and the blue dots have the *same* cardinality, as in the three scenes in Figure 2, iff the yellow dots correspond one-to-one with the blue dots.



a b c

Figure 2

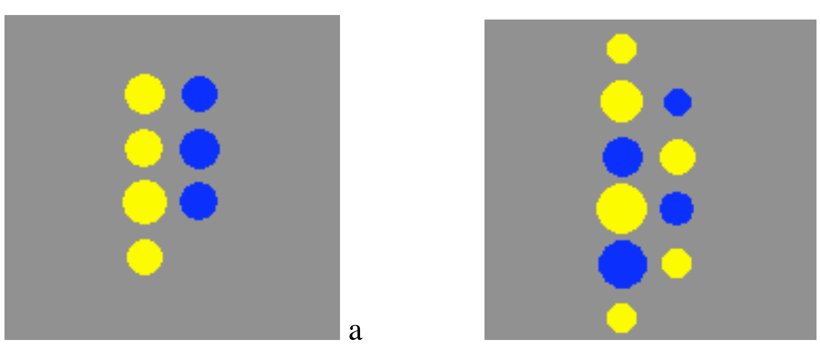
Put another way, the yellow dots and the blue dots have the same cardinality iff there is a function **F** such that: for each yellow dot x, there is a blue dot y, such that **F**(x) = y & **F**(y) = x. And this truism has a consequence worth noting.

A thinker might be able to determine that some things correspond one-to-one with some other things, and hence that the former are *equinumerous* with the latter, even if the thinker is unable to determine the shared cardinality in question. Consider, for example, Figure 2b. One need not know that there are four yellow dots, and four blue dots, in order to know that there are (exactly) as many yellow dots as blue dots. We will return to the relevant generalization, often called "Hume's Principle," which lies at the heart of arithmetic.

$$\text{(HP)}\;\; \#\{x: \Delta(x)\} = \#\{x: \Psi(x)\} \;\text{ iff }\; \text{OneToOne}[\{x: \Delta(x)\}, \{x: \Psi(x)\}]$$

Tacit knowledge of this generalization, ranging over predicates Δ and Ψ, may play an important role in mature numerical competence.[iv] But for now, we just want to note that recognizing *equi*numerosity does not require a capacity to recognize, compare, and identify cardinalities.

Correlatively, a thinker might be able to recognize *non*equinumerosity—and thereby determine which of two sets has the *greater* cardinality—without having a capacity to recognize, compare, and distinguish cardinalities. This is important, in thinking about what 'most' means. For it reminds one that counting is not the only way of determining whether or not some things outnumber some other things. In each scene in Figure 3, there are more yellow dots than blue dots.



a b

Figure 3

One can see that this is so, and hence that most of the dots are yellow, without counting or otherwise figuring out the number of yellow dots. It suffices to note that *some but not all of* the yellow dots can be put in one-to-one correspondence with *all of* the blue dots.

Put another way, the yellow dots outnumber the blue dots—the set of yellows has a greater cardinality than the set of blues—iff some proper subset of the yellow dots has elements that correspond one-to-one with the (elements of the set of) blue dots. So let's define a relation, OneToOne+, as shown below.

$$\text{OneToOne+}[\{x: \Delta(x)\}, \{x: \Psi(x)\}] \text{ iff}$$

$$\text{for some set } \mathbf{s}, \mathbf{s} \subseteq \{x: \Delta(x)\} \ \& \ \text{OneToOne}[\mathbf{s}, \{x: \Psi(x)\}]$$

Then for cases involving finitely many things, the following generalization is correct.

$$\#\{x: \Delta(x)\} > \#\{x: \Psi(x)\} \text{ iff } \text{OneToOne+}[\{x: \Delta(x)\}, \{x: \Psi(x)\}]$$

That is, the number of the Δs is greater than the number of the Ψs iff some (but not all) of the Δs correspond one-to-one with (all of) the Ψs. So as a special case, in any given domain, the Ψs outnumber the non-Ψs iff some of the Ψs correspond one-to-one with all of the non-Ψs.

$$\#\{x: \Psi(x)\} > \#\{x: \neg\Psi(x)\} \text{ iff } \text{OneToOne+}[\{x: \Psi(x)\}, \{x: \neg\Psi(x)\}]$$

So most dots are yellow iff OneToOne+[{x: Dot(x) & Yellow(x)}, {x: Dot(x) & ¬Yellow(x)}]. Thus, (5) is true iff some of the yellow dots correspond one-to-one with all the nonyellow dots.

(5) Most dots are yellow

With this in mind, it seems rash to insist that (5) *means* that the number of yellow dots exceeds the number of nonyellow dots. Though neither should one insist that (5) is understood in terms of one-to-one correspondence. Given formally distinct but truth-conditionally equivalent ways of describing the meaning of 'most', one wants to know if there is a corresponding psychological distinction—and if so, how to characterize it in a theoretically useful way.

One possibility is that there is no fact of the matter to discern, in much the way that there is no fact of the matter about whether temperatures should be measured in Farenheit or Centigrade. Perhaps describing the meaning of (5) in terms of cardinalities is just as good, and just as bad, as describing that meaning in terms of one-to-one correspondence. Some distinctions are merely notational, or at least not psychologically significant. But another possibility is that the contrasting formal descriptions correspond, in some way that needs to be made more explicit, with contrasting psychological hypotheses about how 'most' is understood.

Of course, one wants reasons for thinking there is a contrast for the formal distinction to reflect. But as discussed below, competent speakers can reliably answer questions involving 'most' in situations that preclude counting, thus raising questions about any hypothesized meaning for 'most' based on exact cardinalities. Moreover, in other work, we have found that some children who do not yet have exact cardinality concepts still seem to understand 'most' appropriately (Halberda, Taing and Lidz, 2007). These facts are not decisive; participants might understand 'most' in terms of cardinalities, and still have other methods of evaluating most-sentences for truth or falsity, when counting is impossible or not worth the bother. Nonetheless, an important point remains: formal differences can at least suggest different hypotheses about the *algorithms* that competent speakers might use to evaluate sentences containing 'most'.

There can be no guarantee that a single algorithm is used across individuals in any one context, or even that any one individual understands 'most' the same way across contexts. But it may be that certain algorithms, including some that initially seem like plausible candidates, are not used in situations that positively invite their use. This is one way in which formal distinctions can be translated into competing psychological hypotheses, and comparing such hypotheses can inform semantic theorizing. If there is no psychological evidence of speakers using a

OneToOne+ strategy, in situations constructed to be favorable for such use, this militates against specifying the meaning of 'most' in terms of OneToOne+.[v]

This is the kind of conclusion we will urge, in opposition to the initially attractive idea that speakers understand 'most' in terms of one-to-one correspondence. But it doesn't follow that speakers understand 'most' in terms of cardinalities. (There are still more characterizations of the relevant lexical meaning to consider.) Thus far, we have not been trying to advance a specific proposal. Rather, we wanted to focus attention on the relation—or perhaps lack of relation—between current semantic theories and specific psychological hypotheses about how competent speakers understand linguistic expressions.[vi]

Summarizing, there can be many truth-conditionally equivalent ways of characterizing the contribution of 'most' to the truth conditions of declarative sentences in which the word appears. Each characterization may suggest a conceptual representation corresponding to the lexical item. But given equivalences like (HP), it can be hard to distinguish alternatives.

$$\text{(HP)}\quad \#\{x\text{: }\Delta(x)\} = \#\{x\text{: }\Psi(x)\} \text{ iff } \text{OneToOne}[\{x\text{: }\Delta(x)\}, \{x\text{: }\Psi(x)\}]$$

Still, we think that progress is possible.

2. Getting Past Level One

Imagine three people, each with their own definition of 'most'. Alex learned 'most' explicitly in terms of cardinalities and the arithmetic relation greater-than, while Bobby (who cannot count) has always defined 'most' in terms of one-to-one correspondence. Nonetheless, Alex and Bobby always agree about whether sentences of the form 'Most Δs are Ψ' are true. Chris is different. Suppose, to foreshadow discussion in section three, that Chris has an "Approximate Number System," which interacts with other cognitive mechanisms to generate 'most'-judgments as follows: in cases that are *not* close calls, say 9 yellow dots versus 4 blue dots, Chris shares the judgments of Alex and Bobby; but in close cases, say 8 versus 7, Chris cannot tell the difference (and defers to others). Alex, Bobby, and Chris are able to communicate well enough for most purposes. In this sense, they understand each other. But there is an important sense in which their idiolects and underlying psychologies differ; see note six.

It is useful, in this context, to draw a distinction similar to Marr's (1982) contrast between Level One (computational) and Level Two (algorithmic) questions. Given a system that seems to be performing computations of some kind, we can distinguish Level One questions about *what* the system is computing from Level Two questions about *how* the system is computing it. With regard to understanding sentences, this distinction has a familiar application that can be extended to understanding words.

Given some assumptions about the kinds of sentential properties that competent speakers can recognize—say, the truth conditions of novel declarative sentences—one can ask *how* speakers recognize these properties. In practice, semanticists often stop short of offering answers, even when they specify algorithms that determine truth conditions on the basis of hypothesized properties of words and modes of grammatical combination. Such algorithms are rarely put forward as explicit Level Two hypotheses.[vii] Often, it is hard enough to find *one* good way of characterizing the "composition function" that speakers somehow compute, and even harder to find evidence that would favor a specific "composition algorithm." But it is widely agreed that complex expressions are understood, one way or another, by employing algorithms that are compatible with constraints imposed by natural language syntax. And we think that much of semantics should be viewed as an attempt to describe certain properties of expressions *in order* to describe the algorithms that speakers employ in determining those properties compositionally.

Similarly, one can and should distinguish the truth conditional contribution of a word like 'most' from how that contribution is represented in the minds of competent speakers. It is useful, and perhaps essential, to begin with a proposal about what the truth conditional contribution is. For these initial purposes, the many truth conditionally equivalent descriptions will be equivalent. If the task is to describe the distinctive contributions of determiners like 'every' and 'most'—in a way that permits formulation of *some* algorithm that determines the truth conditions of relevant sentences, given syntactic structures and lexical specifications—then theorists have wide latitude with regard to how they specify the truth conditional contributions of the determiners. This task is not trivial. But depending on the proposed method of computation, it might be enough to say that each determiner indicates a certain function from pairs of predicates to truth conditions (see note two); in which case, endlessly many formally distinct lexical specifications will be acceptable. On other hand, if the task is to say how speakers understand 'most'—with the aim of saying how linguistic comprehension frames the task of evaluating sentences of the form 'Most Δs are Ψs'—then one must get beyond Level One questions about which function 'most' indicates.

It may not be possible, at this early stage of inquiry, to formulate defensible "Level 2" hypotheses about exactly how speakers represent and compute any particular function. Still, one can try to formulate and defend "Level 1.5" hypotheses about the *kinds* of representations that children and adults employ in understanding the word 'most'.[viii] As discussed in section one, the following representations are truth conditionally equivalent, given finitely many things.

$$\#\{x: Dot(x) \;\&\; Yellow(x)\} > \#\{x: Dot(x) \;\&\; \neg Yellow(x)\}$$

$$OneToOne{+}[\{x: Dot(x) \;\&\; Yellow(x)\}\;,\; \{x: Dot(x) \;\&\; \neg Yellow(x)\}]$$

But only the first makes reference to cardinalities and a relation between them. This can be made explicit, by replacing the first representation with (5a), which is equivalent for our purposes here.

(5a) GreaterThan[#{x: Dot(x) & Yellow(x)}, #{x: Dot(x) & ¬Yellow(x)}]

(5b) OneToOne+[{x: Dot(x) & Yellow(x)}, {x: Dot(x) & ¬Yellow(x)}]

Of course, (5a) can itself be rewritten in many ways, as can (5b). Nonetheless, the formal distinction between (5a) and (5b) suggests two different *classes* of algorithms for determining whether or not most dots are yellow: algorithms that involve representing two cardinalities, and determining whether the first is greater than the second; and algorithms that test for a certain kind of correspondence relation, between (the elements of) two sets, without ever representing the cardinalities of those sets. This leaves room for various strategies in making the relevant determinations. The number of yellow dots might be determined by counting, or by recognizing a pattern already associated with a number. Counting might start with zero or one. A mechanism might determine whether one number is bigger than another by consulting memory, or by subtracting and checking for a (positive) remainder.[ix] There are likewise many ways of determining whether some things correspond one-to-one with some other things, and whether such a correspondence leaves any "outliers." The central issue at present is that there be *some* psychologically plausible way of assessing the relevant inputs and computing the functions for these two classes of algorithms: the first relying on comparison of cardinalities, and the second relying on detection of one-to-one correspondence (and a remainder).

For several reasons, psychological and semantic/logical, one might initially bet on the latter (OneToOne+) family of algorithms—at least as an evaluation procedure, and perhaps as a way of understanding 'most'. Psychological considerations suggest that counting is often slower and more laborious than evaluating sentences with 'most'. In situations where a subject *cannot* count yet the subject can still determine whether or not most Δs are Ψs, as when items are briefly flashed, appealing to a OneToOne+ algorithm might seem especially attractive. And as noted above, some children who have not yet learned how to count can still evaluate whether or not most of the Δs are Ψs (Halberda, Taing & Lidz, 2007). One might think that such children understand 'most' in terms of OneToOne+. In which case, parsimony would suggest that even participants who can count still *understand* 'most' in these terms.

With respect to semantics/logic, recall the generalization (HP),

$$\text{(HP)}\ \ \#\{x: \Delta(x)\} = \#\{x: \Psi(x)\}\ \text{ iff }\ \text{OneToOne}[\{x: \Delta(x)\}, \{x: \Psi(x)\}]$$

which reflects the deep relation between counting and one-to-one correspondence. It turns out that all of arithmetic follows from (HP), given a consistent logic that is presupposed by any plausible semantics for natural language. This remarkable fact, essentially proved by Frege (1884, 1893, 1903), illustrates the power of the modern logic that Frege (1879) invented and contemporary semanticists regularly employ.[x] Given this logic, (HP) encapsulates arithmetic. The apparent simplicity of one-to-one correspondence, and its relation to the foundations of arithmetic, makes it tempting to think that 'most' (along with other counting/cardinality expressions) is understood in terms of a basic concept of one-to-one correspondence. Indeed, one might think of counting as a laborious but accurate way of determining correspondence relations.

So if the task is to determine whether or not most dots are yellow, one might expect speakers to jump at the chance of answering *without* having to determine the number of yellow dots (and nonyellow dots). Anecdotally, this was the expectation of most semanticists and logicians we queried, including one or more authors of this paper. At a minimum, a OneToOne+ strategy would seem to be a plausible candidate for speakers who do not know how to count, or for numerate speakers in situations where dots are presented too quickly for counting. But our experiment, described in section five, suggests that this strategy is not available to participants.

In the experiment, we vary the spatial arrangements and cardinalities of yellow dots and nonyellow dots in a search for evidence of participants using a OneToOne+ algorithm to determine whether or not most of the dots are yellow. If no such algorithm is used, even in situations that positively invite its use, then it is (we assume) unlikely that participants understand 'most' in terms of one-to-one correspondence. Moreover, we offer positive evidence that participants used representations that do not support a OneToOne+ strategy. This tells even more forcefully against OneToOne+ hypotheses about how 'most' is understood.[xi]

3. Background Psychology, Including The ANS

The two classes of algorithms we have been discussing, corresponding to (5a) and (5b),

(5a) GreaterThan[#{x: Dot(x) & Yellow(x)}, #{x: Dot(x) & ¬Yellow(x)}]

(5b) OneToOne+[{x: Dot(x) & Yellow(x)}, {x: Dot(x) & ¬Yellow(x)}]

involve two kinds of representational resources for which there is evidence in the psychological literature. First, there are representations of exact cardinal values—available to adults who know the meanings of words like 'seven', 'nine' and 'sixty'—that might support a meaning for 'most' based on comparing cardinalities.[xii] Second, demonstrations of object-tracking competence in infants (Wynn, 1992; Feigenson, 2005) have revealed a cognitive system that can detect one-to-one correspondences, at least in certain situations; and this system might support a meaning based on OneToOne+. In this sense, both "Cardinality-most" and "Correspondence-most" are viable hypotheses on psychological grounds. But it is also useful to consider a third representational resource, the Approximate Number System, that might provide numerical content for the meaning of 'most'. This content, when applied to representing linguistic quantifiers, may or may not be dependent on an appreciation of cardinalities. For this reason we treat it separately from Cardinality-most and Correspondence-most.

Before learning how to count, children have an approximate sense of the number of items in an array. Like many nonverbal animals, including rats and pigeons, human infants have an Approximate Number System (ANS): an evolutionarily ancient cognitive resource that generates representations of numerosity across multiple modalities (e.g. for sets of visual objects, auditory beeps, and events such as jumps, presented either serially or in parallel). The ANS does not require explicit training with numerosity in order to develop, and the brain areas that support this system in humans and in other primates have been identified (for review see Feigenson, Dehaene & Spelke, 2004). "The ANS generates representations of pluralities in ways that effectively order those pluralities according to cardinality —albeit stochastically, and within certain limits described by Weber's Law (for review see Dehaene, 1997)."

Weber's Law, which applies to many kinds of representation (e.g. loudness, weight, brightness), states that discriminability depends on the ratio of relevant representational values. With respect to the ANS, 6 things are detectably different from 12 to the same extent that 60 things are detectably different from 120. In each case, the Weber Ratio is 2 (WR = larger set #/ smaller set #). If the absolute numeric difference between the comparison groups is maintained, but the numerosities are increased (e.g. from 6 vs 12 to 12 vs 18, with a constant difference of 6), discriminability will become poorer. This is the so-called size effect. There is also a distance effect. If the cardinality of one group is held constant while the other changes (e.g. from 6 vs 12 to 6 vs 18), discriminability increases with greater numeric distance between the cardinalities. Representations of the ANS also seem to be integrated with adult understanding of exact cardinalities, since reactions to questions that seem to be about cardinalities or numerals—e.g., ‘Is 67 bigger than 59’—also exhibit size and distance effects. (For review see Dehaene, 1997.)

It will be useful to consider a certain model of the representations generated by the ANS. It is generally agreed that each numerosity is mentally represented by a distribution of activation on an internal “number line” constituted by a range of possible ANS-representations that exhibit certain global properties. The distributions in question are inherently noisy, and they do not represent number exactly or discretely (e.g., Dehaene, 1997; Gallistel & Gelman, 2000). The mental number line is often characterized as having linearly increasing means and linearly increasing standard deviation (Gallistel & Gelman, 2000), as in Figure 4. In this figure, numerosities are represented with Gaussian curves, with the discriminability of any two numerosities being a function of the overlap of the corresponding Gaussians: the more overlap, the poorer the discriminability. For example, the curves corresponding to 8 and 10 overlap more than the curves corresponding to 2 and 4. And notice that all the representations, even the first one, are Gaussian curves. In this model, the ANS has no discrete representation of unity; no curve represents one (or more) things as having a cardinality of exactly one.[xiii] We return to this point.

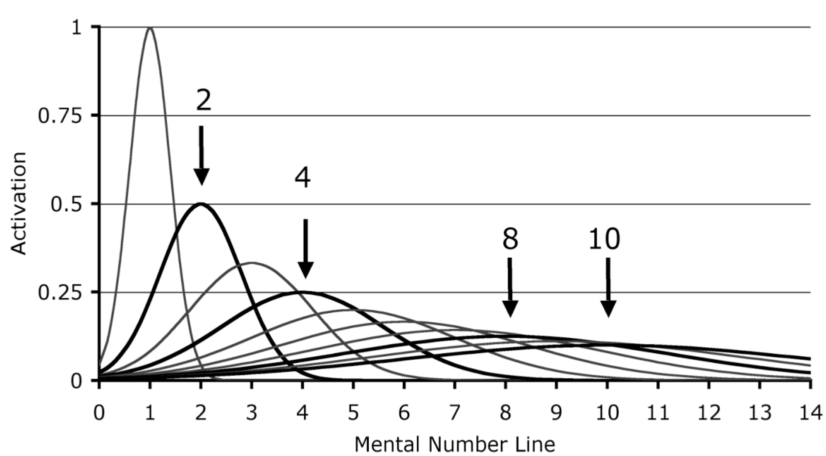


Figure 4

The acuity of the ANS improves during childhood. In terms of the model, the spread of Gaussian curves decreases with age. Adults can discriminate numerosities that differ by at least a 7:8 ratio (Halberda & Feigenson, 2007; Barth, 2003; van Oeffelen & Vos, 1982). But for 6 month old infants, numerosities must differ by at least a 1:2 ratio in order for discrimination to be accurate (Xu & Spelke, 2000). In both human adults (Halberda, Sires & Feigenson, 2006) and infants (Zosh, Halberda & Feigenson, 2007), the ANS is capable of generating numerosity estimates for up to three sets in parallel—enough for the apparent numerical content of a determiner like 'most', which might compare two cardinalities.[xiv] The capacity for building multi-set representations in infants suggests the potential relevance of the ANS for supporting comparative determiners at the earliest ages of language learning.

By 5 years of age, in numerate cultures, representations of the ANS have been mapped onto the discrete number words (LeCorre & Carey, in press). The ANS is activated anytime a numerate adult sees an Arabic numeral, hears or reads a number word or performs a mental operation on numbers such as subtraction (for review see Dehaene, 1997). The ANS generates representations of numerosity very rapidly. Imagine flashing an array of items (e.g. 25 yellow dots) for only 250 ms, too fast for explicit counting. The recording of single neurons—in the physiological analog of the ANS, in awake behaving monkeys who are shown such arrays—suggests that the ANS can generate a representation of the approximate number of items present within 150 ms of stimulus onset. That is, a representation/Gaussian of *approximately 25* can be generated with almost incredible speed (Nieder & Miller, 2004). When shown such arrays, adults and children over 5 can produce a discrete numerical estimate (e.g. 'about 20 yellow dots'). Over many trials, the pattern of the numerical estimates given by adults and children will follow the same shape as the Gaussian curves in Figure 4. For example, when shown many instances of 25 dots, ‘twenty-five’ will be the most common answer; though participants will also sometimes say ‘twenty-four’ or ‘twenty-six’, ‘twenty’ or ‘thirty’ (etc.), with the probability of saying a given number word being a smooth Gaussian curve centered on ‘twenty-five’. (Halberda et al, 2006; LeCorre & Carey, in press; Whalen, Gallistel & Gelman, 1999). This demonstrates that the representations of the ANS can be mapped to discrete number words, and thereby discrete cardinal values, albeit in a noisy and approximate way.

Consider how the representations of the ANS might be relevant to the meaning of ‘most’. As noted above, the ANS will not deliver a representation of discretely and exactly one item. If only for this reason, speakers cannot rely on the representations of the ANS to implement a OneToOne+ algorithm (for understanding or evaluation). But representations of the ANS do represent something like cardinalities as stochastically ordered pluralities, and these representations can be mapped onto discrete number words. This suggests two possible ways of using ANS-representations: first, implement a meaning for ‘most’ (an algorithm for understanding ‘most’-sentences) via the stochastic representations of the ANS itself, without reference to any discrete cardinalities; or second, rely on the ANS to evaluate ‘most’-sentences (with ‘most’ having the meaning of ‘Cardinality-most’) via the mapping between the stochastic ANS values and discrete number concepts.

The first option might be exploited by noncounters (i.e. participants without access to the discrete exact cardinalities generated via the counting routine), who may not have any concept of cardinality to associate with a word like ‘most’. Children, in particular, may not have thoughts with the structure or truth-conditional precision of (5a).

(5a) GreaterThan[#{x: Dot(x) & Yellow(x)}, #{x: Dot(x) & ¬Yellow(x)}]

Numerate adults may be able to associate 'most' with a (total) function from pairs of predicates to truth or falsity. But other speakers may be like the hypothetical Chris discussed in section two. For these individuals, (the sound of) 'most' may indicate a *partial* function: one that fails to map a pair of predicates $< \Delta, \Psi>$ to truth or falsity when the number of Δs that are Ψs fails to *differ enough* from the number of Δs that are not Ψs.[xv] It is possible that noncounters use a homophone of 'most' that is a "partial variant" of our word. Put another way, our adult word 'most' may be a "cardinal precisification" of an earlier word—call it 'pmost', with silent 'p'—that has the following characteristic: when most of the dots are yellow, and it is not a close call, 'Pmost of the dots are yellow' is true; when it is false that most of the dots are yellow, and it is not a close call, 'Pmost of the dots are yellow' is false; but sometimes, as when 51% or 49% of the dots are yellow, 'Pmost of the dots are yellow' is neither true nor false. In which case, 'pmost' would be similar to 'most', but semantically rooted in the stochastic representations of the ANS.

The second option, for how the ANS might be related to 'most', may be exploited by numerate speakers faced with situations in which it is impossible or inconvenient to determine relevant cardinalities exactly. Even adults who *understand* 'most' in terms of a relation between cardinalities might be able to use their ANS to defeasibly evaluate 'most'-claims. One can imagine a "numeralizing waystation"—perhaps restricted to speakers (over 5 years of age) who have mapped representations/Gaussians of the ANS to discrete number words—that associates ANS-representations with independent representations of exact cardinal values. For example, an ANS-representation that is usually triggered by pluralities of six (though sometimes five or seven) might be associated with a mental analog of '6', while an ANS-representation usually triggered by pluralities of nine (though sometimes eight or ten, and occasionally seven or eleven) might be associated with a mental analog of '9'. Given some such waystation that interfaces between the ANS and other cognitive systems, 'most'-sentences could be *understood* as in (5a) yet regularly (and imperfectly) evaluated via the ANS.

4. Putting All This Together

By way of summarizing to this point, let's stipulate that the Tmost-function (which may be applicable only to the count-noun cases we are considering), is a *total* function that can be characterized in many ways: $>[\#\{x: \Delta(x) \ \& \ \Psi(x)\}, \#\{x: \Delta(x) \ \& \ \neg\Psi(x)\}]$; $\text{OneToOne+}[\{x: \Delta(x) \ \& \ \Psi(x)\}, \{x: \Delta(x) \ \& \ \neg\Psi(x)\}]$; etc. And let's stipulate that the Pmost-function is a (partial) variant that can be characterized in terms of the stochastic representations of the ANS. The possibilities we have outlined are diagrammed in Figure 5 below.

Speakers who associate (the sound of) 'most' with the Tmost-function might understand 'Most of the dots are yellow' in at least two ways: as a claim that logically implies cardinalities (Cardinality-most), or as the less numerically loaded claim that *some* of the yellow dots correspond one-to-one with *all* of the nonyellow dots (Correspondence-most). Cardinalities can be determined by counting. But another strategy is to generate ANS-representations that can be sent to a numeralizing waystation that associates such representations with mental analogs of written numerals. (This would be conceptual machinery similar or identical to that which enables children and adults to produce a discrete number word in response to seeing a quickly flashed array of dots which activates an ANS representation.) Positing such a waystation is indicated, in figure 5, with the box containing '#'. We also note, for completeness, two possibilities that seem unlikely (see notes nine and eleven): given a psychologically realized version of Hume's Principle, indicated with 'ΨHP', a thinker might *understand* 'most' in terms of cardinality but (whenever possible) *evaluate* by checking for OneToOne+ correspondence—or conversely, understand 'most' in terms of OneToOne+ but default to evaluating by counting.[xvi] Finally, and more interestingly, ‘most’ might be associated (at least in noncounters) with the Pmost-function, with both understanding and evaluation having the stochastic character of the ANS.

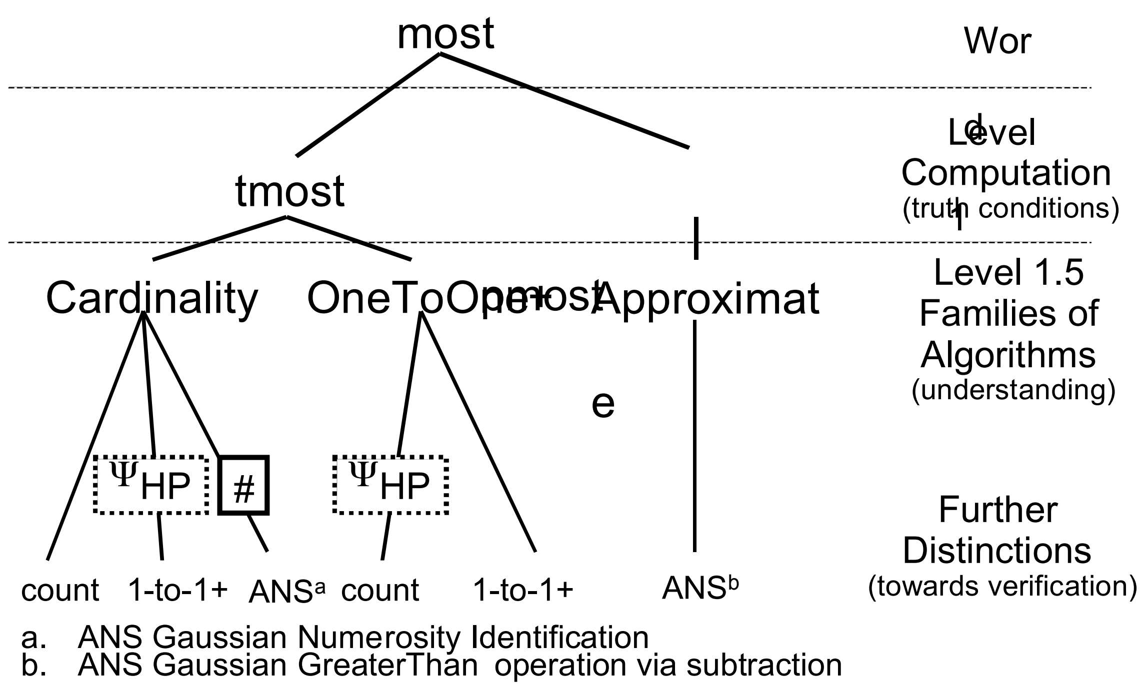


Figure 5

This map of options invites the question of which, if any, can be discredited by experimental methods? We focused on the initially tempting hypothesis that ‘most’ is understood in terms of OneToOne+, in part because ruling out this class of possibilities would be a substantive kind of progress that is achievable with current methods. The basic idea is simple: present participants (at time scales that make counting impossible) with scenes in which it is easy to employ a OneToOne+ strategy, and scenes in which it is hard to employ this strategy. If this variation does not affect participants’ accuracy in evaluating the sentence ‘Most of the dots are yellow’, that tells against the initially tempting hypothesis.

In our experiment, we presented competent adult English speakers with various arrays of yellow and blue dots. On each trial, the subject was required to say whether or not most of the dots were yellow. Arrays were flashed too quickly for exact counting to be possible (i.e. 200 ms), thus precluding any branch in Figure 5 with 'count' as the method of evaluation. Across trials, arrays varied how easy or hard it was to apply a OneToOne+ strategy. A null result of this manipulation argues against OneToOne+ as the underlying meaning of 'most'. More strongly, we looked for positive evidence of participants relying on representations of the ANS. Participants were presented with varying ratios of yellow to blue dots. And so the hypothesis that ANS representations were used, in evaluating 'most'-sentences, leads to a very specific set of predictions (garnered from the literature on the ANS and classic psychophysics) concerning how performance should vary as a function of ratio.

Evidence of ANS use will not, however, distinguish the two possibilities in Figure 5 that list 'ANS' as a means of evaluation: via the ANS through the numeralizing waystation (Tmost); or through an ANS-specific operation, which we will call Gaussian subtraction (Pmost), described in section 5. Deciding between these two options is not our goal in this paper. Rather, the focus is on OneToOne+. And to repeat an earlier point, the ANS does not have a discrete exact representation of a single item; hence, it cannot be used to check for one-to-one correspondences. So any positive evidence of the use of ANS representations in our experiment constitutes strong evidence against the hypothesis that the function associated with 'most' (Tmost) is computed in terms of a OneToOne+ algorithm.

5. Experiment

We used a common visual identification paradigm to test how speakers understand 'most'.

Method

Participants

Twelve naive adults with normal vision each received $5 for participation.

Materials and Apparatus

Each participant viewed 360 trials on an LCD screen (27.3 X 33.7 cm). Viewing distance was unconstrained, but averaged approximately 50 cm. The diameter of a typical dot subtended approximately 1 degree of visual angle from a viewing distance of 50 cm.

Design and Procedure

On each trial, participants saw a 200ms display containing dots of two colors (yellow and blue). Participants were asked to answer the question 'Are most of the dots yellow?' for each trial. The number of dots of each color varied between five and seventeen. Whether the yellow set or the blue set was larger (and hence, whether the correct answer was 'yes' or 'no') was randomized. Participants answered 'yes' or 'no' by pressing buttons on a keyboard.

Each trial came from one of nine "bins", each characterised by a ratio. The first bin contained trials where the ratio of the smaller set to the larger set was close to 1:2; the second bin contained trials where the ratio was close to 2:3; and the remaining bins contained trials close to 3:4, 4:5, ..., 9:10. Each participant received ten trials in each bin for each of four conditions: Scattered Random, Scattered Pairs, Column Pairs Mixed and Column Pairs Sorted. The total number of trials for each participant was therefore 9 ratios x 4 conditions x 10 trials = 360. These were presented in randomized order.

On Scattered Random trials, all the dots (blue and yellow) were scattered randomly throughout the display. See Figure 6a. In each of the other three conditions, dots were displayed in some way intuitively amenable to a "one-to-one pair off" algorithm (OneToOne+), with yellow dots and blue dots occurring in pairs. On Scattered Pairs trials, every dot from the smaller set was displayed paired with (approximately four pixels away from) a dot from the larger set;

the remaining dots from the larger set were scattered randomly. See Figure 6b. On Column Pairs Mixed trials, dots were arranged in a grid with two columns and n rows, where n is the size of the larger set. Each row had either one dot from each set, or a single dot from the larger set with the position (left column or right column) for each item being determined randomly for each row. See Figure 6c. On Column Pairs Sorted trials, dots were likewise arranged in two columns and n rows, but with all the yellow dots in one column and all the blue dots in the other. The smaller set of dots was grouped together from the top of its column, with no empty rows between dots, so that the display consisted essentially of two parallel lines of dots with side (yellow on left column or right column) determined randomly. See Figure 6d.

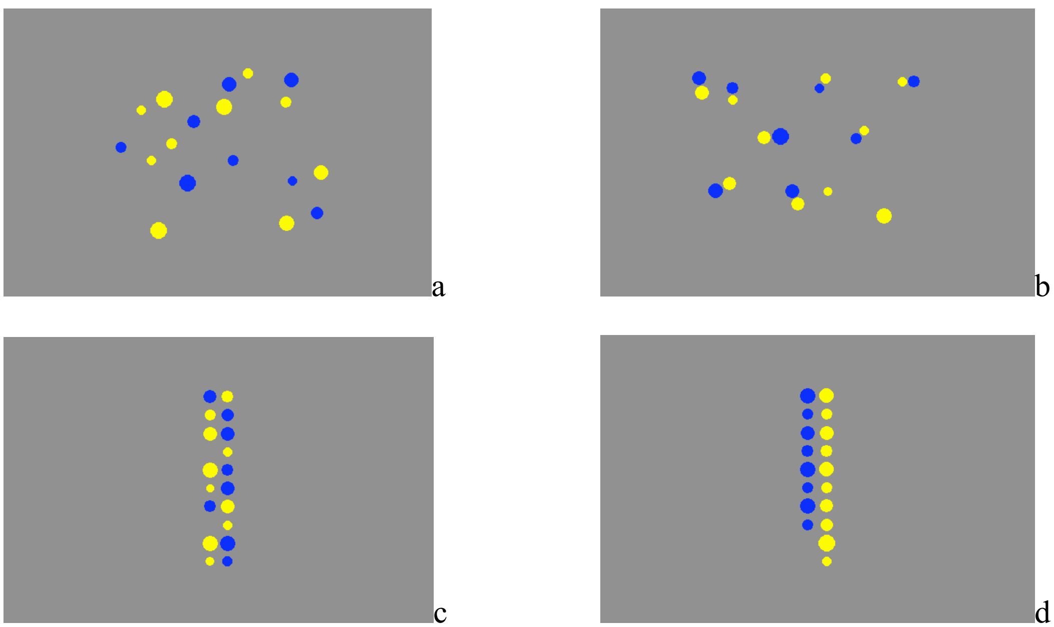


Figure 6

Half of the trials for each trial type for each ratio were "size-controlled:" while individual dot sizes varied, the size of the average blue dot was equal to the size of the average yellow dot, so the set with more dots would also have a larger total area on the screen (i.e. more blue pixels when more dots were blue). The other half of the trials were "area-controlled:" individual dot sizes varied, but the number of blue pixels was also the number of yellow pixels (i.e., smaller blue dots on average when more dots were blue). On both size-controlled and area-controlled trials, individual dot sizes were randomly varied by up to 35% of the set average. This discouraged the use of individual dot size as a proxy for number.

Results

Percent correct for each participant for each ratio was entered into a 4 Trial Type (Scattered Random, Scattered Pairs, Column Pairs Mixed, Column Pairs Sorted) X 2 Stimulus Type (size-controlled, area-controlled) X 9 Ratio Repeated Measures ANOVA. There was a significant effect of Ratio, as participants did better with easier ratios: $F(8, 72) = 13.811$, $p < .001$; a significant effect of Trial Type, as participants did better on Column Pairs Sorted trials: $F(3, 27) = 47.016$, $p < .001$; no effect of Stimulus Type, as participants did equally well on size-controlled and area-controlled trials: $F(1, 9) = 1.341$, $p = .277$; and a marginal Trial Type X Ratio interaction, as participants did better on difficult ratios on Column Pairs Sorted trials: $F(24, 216) = 1.432$, $p = .094$. Participants did equally well on size-controlled and area-controlled trials, indicating that they relied on the *number* of dots and not continuous variables such as area that are often confounded with number. Performance for each participant for each ratio was combined across Trial Type for further analyses.

Planned Repeated Measures ANOVAs compared performance pair-wise for each Trial Type. Performance on Scattered Random, Scattered Pairs, and Column Pairs Mixed all patterned together with no significant differences, whereas performance on each of these conditions was significantly worse than that on Column Pairs Sorted trials. The F and *p* values for these comparisons are listed in Table 1. This pattern can also be seen in Figure 7. Contrary to what would be expected from use of a OneToOne+ algorithm, performance on Scattered Random trials patterned with performance on Scattered Pairs and Column Pairs Mixed, with percent correct declining as a function of Weber Ratio (# of larger set/ # of smaller set). Performance on Column Pairs Sorted trials remained at ceiling for all ratios tested, suggesting that a different process was used to verify 'most' on these trials.

Table 1. Pairwise comparison of trial types

Trial Types	F	*p*
Scattered Random-Scattered Pairs	.216	.651
Scattered Random-Column Pairs Mixed	.446	.518
Scattered Pairs-Column Pairs Mixed	.127	.728
Column Pairs Sorted- Scattered Random	152.17	.0001
Column Pairs Sorted- Scattered Pairs	193.89	.0001
Column Pairs Sorted- Column Pairs Mixed	131.66	.0001

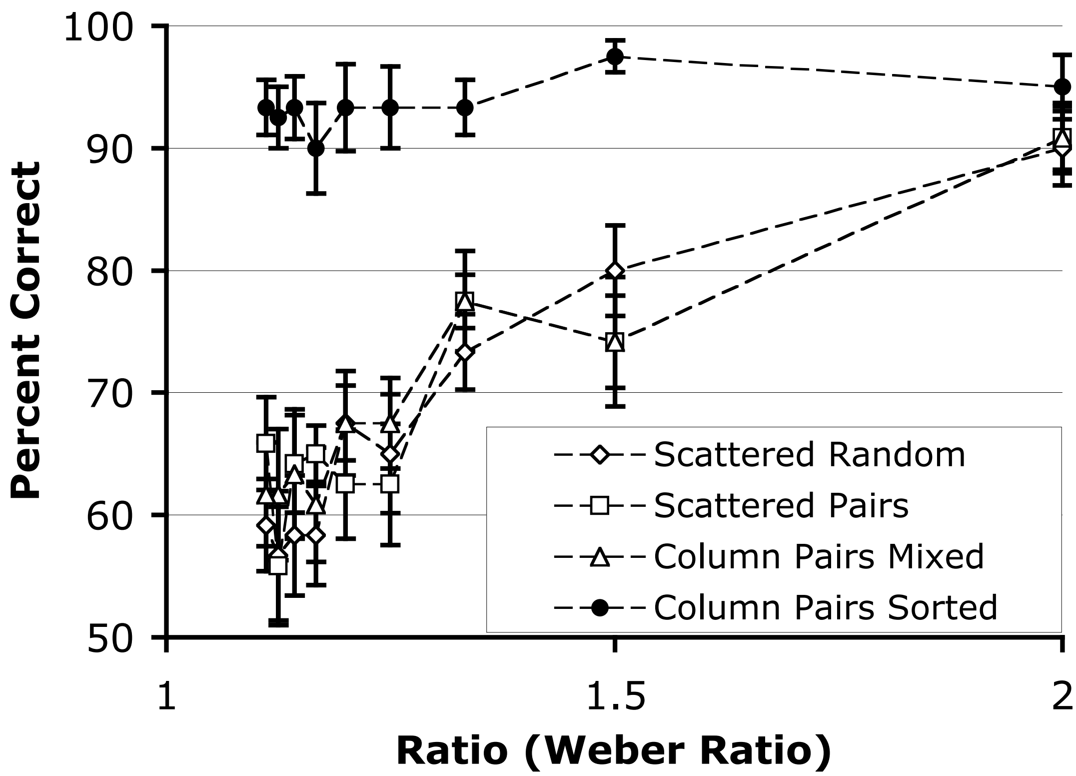


Figure 7

If participants relied on the representations of the Approximate Number System to verify 'most' on Scattered Random, Scattered Pairs, and Column Pairs Mixed trials, performance on these trials should accord with a model of the psychophysics of this system. We rely on a classic psychophysical model that has been used by labs other than our own, indicating its acceptance in the literature (e.g., Pica et al., 2004). The average percent correct at each ratio across participants is modeled for each Trial Type as a function of increasing Weber Ratio (larger set/smaller set, or n2/n1). Pairs of numerosities are represented as Gaussian random variables X2 and X1, with means n2 and n1, and standard deviations equal to the critical Weber fraction (w) * n. Subtracting the Gaussian for the smaller set from the larger returns a new Gaussian, with a mean of n2-n1 and a standard deviation of $w\sqrt{n1^2+n2^2}$ (simply the difference of two Gaussian random variables). Correlatively, subtracting X2 from X1 returns a new Gaussian random variable that has a mean of n1-n2, and percent *correct* can be calculated from this Gaussian as the area under the curve that falls to the right of zero, computed as:

$$\frac{1}{2}\ \mathrm{erfc}\left(\frac{n_1 - n_2}{\sqrt{2}w\sqrt{n_1^2 + n_2^2}}\right) \times 100$$

The one free parameter in this equation is the critical Weber Fraction (w). This parameter determines percent correct for every Weber Ratio (n2/n1). The mean of subject means for percent correct at each of the nine ratio bins and the theoretically determined origin of the function (50% correct at Weber Ratio = 1, where the number of blue dots and yellow dots would in fact be identical), were fit using this psychophysical model. As can be seen in Figure 8, the fits for Scattered Random, Scattered Pairs, and Column Pairs Mixed trials fell directly on top of one another. Table 2 summarizes the R^2 values, the estimated critical Weber fraction, and the nearest whole-number translation of this fraction for each fit. These R^2 values suggest agreement between the psychophysical model of the ANS and participants' performance in the most-task (R^2 values > .85). The critical Weber fraction on these trial types suggests that participants relied on the representations of the Approximate Number System to evaluate 'most'.

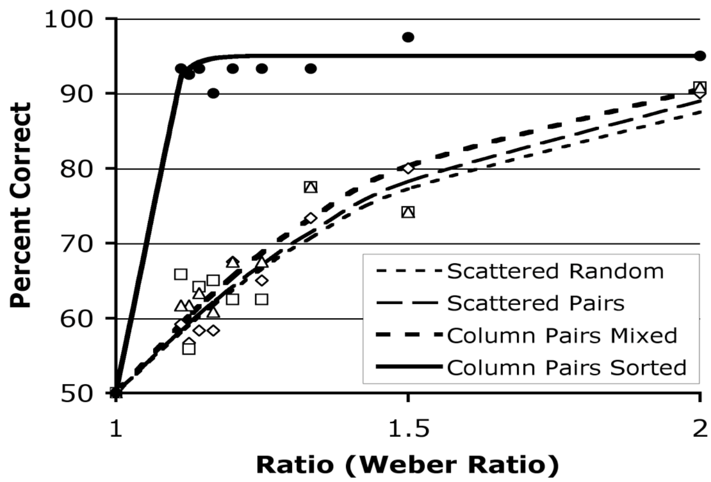


Figure 8

Table 2. Parameter estimates from psychophysical model

Trial Type	R^2	Critical Weber Fraction	Nearest Whole-Number Ratio
Scattered Random	.9677	.32	3:4
Scattered Pairs	.8642	.33	3:4
Column Pairs Mixed	.9364	.30	3:4
Column Pairs Sorted	.9806	.04	25:26

The critical Weber fraction is expected to be approximately .14 for adults in number discrimination tasks and ‘more’-tasks (Pica et al, 2005), and to range from .14- to .35 in adults when participants are translating ANS representations into whole-number values (via a “numeralizing waystation”), measured as the coefficient of variance (Halberda, Sires & Feigenson, 2006; Whalen et al, 1999). Our estimate of a critical Weber fraction of approximately .3 for three trial types suggests that participants may be translating the representations of the ANS into whole number values via some numeralizing waystation before evaluating ‘most’-sentences. For example, when shown an array of 12 blue and 16 yellow dots, these subsets may activate corresponding ANS representations of numerosity; and these values may be translated into cardinal-number estimates, like ‘twelve’ and ‘sixteen’, for purposes of evaluation—in accord with the underlying concept being *tmost*, with ANS verification. Further work is needed to determine if this is the case, or if adults are relying on the ANS to evaluate *pmost* (see Figure 5).

For comparison, we also fit the data from Column Pairs Sorted trials using the same model of the psychophysics of the ANS. This is to allow a direct comparison to the other 3 trial types, and not to suggest that participants are actually relying on ANS representations on Column Pairs Sorted trials. As can be seen in Figure 8 and Table 2, this model returned a radically different fit for these data, suggesting a critical Weber fraction of .04 or a whole number ratio of 25:26. While participants rely on the representations of the ANS on Scattered Random, Scattered Pairs, and Column Pairs Mixed trials, performance on Column Pairs Sorted trials suggests a different process altogether. In fact, the estimated Weber fraction on Column Pairs Sorted trials (Weber fraction = .04) is very similar to estimates from the literature for human adults detecting the longer of two line segments (Weber fraction = .03) (Coren, Ward & Enns, 1994). Displays for our Column Pairs Sorted trials were constructed such that there is a perfect correlation between the number of dots per subset and the overall length of the column. This means that participants could attend only the length of the column to reach their decision, ignoring however many dots it took to make the column, and translate their judgement of “longer blue column” into a “more blue dots than yellow dots” answer without error. So, even on our Column Pairs Sorted trials, adults do not appear to be engaging in a OneToOne+ procedure. Rather, they are using column length as a proxy for number (the perceptual “lines” created by the blue and yellow columns).

We manipulated the ease of applying a OneToOne+ strategy across trial type and found no effect of this manipulation. But we found evidence that participants relied on the representations of the ANS, representations that are not consistent with a OneToOne+ meaning for ‘most’. These results suggest that the (numerate) adult meaning for ‘most’ is not based on one-to-one correspondence. Before closing the results section, we consider two further issues. First, might an error prone OneToOne+ algorithm be able to capture the pattern of our results? And second, do English speaking adults understand ‘Most of the dots are yellow’ as consistent with ‘There is *one more* yellow than nonyellow dot’, or does the ‘most’-claim imply that there are *significantly more* yellow dots than nonyellow dots?

Some researchers might consider the possibility that our participants were using a OneToOne+ algorithm, but that this algorithm is not responsive to our manipulations in attempts to make OneToOne+ easier (e.g. Scattered Pairs easier than Scattered Random). This hypothesis, however, leaves unexplained why performance declines as a function of decreasing Weber Ratio (i.e. as the number of yellow dots and blue dots becomes closer performance declines). An errorless OneToOne+ algorithm would not show this decrement in performance, but would instead be perfectly accurate so long as there was at least one more yellow than nonyellow dot. To account for this result, one might suggest that the OneToOne+ algorithm is prone to performance errors. For example, an error prone OneToOne+ algorithm might either accidentally use an individual dot to cancel out more than one competitor (e.g. using a dot twice) or fail to use a particular dot altogether. For this error prone OneToOne+ algorithm, as the yellow dots and blue dots become closer in number, such errors would become increasingly meddlesome leading to decreasing accuracy. While this sort of error prone OneToOne+ algorithm might successfully predict decreasing performance as a function of decreasing Weber Ratio, as we shall see, it would still fail to account for a subtlety in the present data.

For any particular trial included in (say) the 3:4 ratio bin, the number of items included for a single color could vary from as few as 5 to as many as 16 items, so that one trial might include 6 yellow and 8 blue dots while another might contain 12 yellow and 16 blue dots. An error prone OneToOne+ algorithm predicts a difference in performance between these two trials. Because the two possible errors described above for an error prone OneToOne+ algorithm work in opposite directions (i.e. using a particular dot more than once and failing to use a particular dot at all), these errors would tend to cancel one another out stochastically, as the total number of dots in a display increases. That is, performance should be better on a 12 yellow versus 16 blue dots trial than on a 6 yellow versus 8 blue dots trial. More generally, use of an error prone OneToOne+ procedure predicts an effect of increasing accuracy with increasing number of dots in the display when ratio is held constant. We tested this prediction with a linear regression on subject means, with percent correct as the dependent variable, and Ratio and Total number of dots in the display as dependent variables. This analysis allows us to ask the following question: while controlling for any possible effects of Ratio, is there any evidence that participants did better as the Total number of dots in the display increased? Contrary to the predictions of an error prone OneToOne+ algorithm we found a marginally significant result in the *opposite* direction: t (367) = -1.952, p = .052, slope = -.365. Participants did slightly *worse* within each Ratio as the Total number of items in the display increased. This analysis shows no support for an error prone OneToOne+ algorithm, and remains consistent with the hypothesis that participants relied on the stochastic representations of the ANS—especially given considerations about possible guessing in participants, or perceptual crowding, which might lead to the slight negative slope[xvii].

A second consideration is that participants might understand ‘Most of the dots are yellow’ as implying that the number of yellow dots is *significantly* greater than the number of nonyellow dots. Instead of the 'tmost' meaning discussed above, which requires only one extra yellow dot, perhaps 'most' imposes the more demanding requirement that there be significantly more yellow dots (by some measure of significance). On this view, a situation with 10 yellow dots and 9 blue dots (≈ 52% yellow) might *not* make 'Most of the dots are yellow' true, even if situations with 14 yellow dots and 9 blue dots (≈ 60% yellow) would. This hypothesis predicts a decrease in participants' willingness to assent to 'Most of the dots are yellow' as Weber Ratio decreases (i.e. as the ratio of yellow to nonyellow dots moves closer to 1:1), independent of whether the more demanding meaning is rooted in cardinalities or in OneToOne+. If participants understood 'most' in this way, our observed result of a decrease in performance with decreasing Weber Ratio could not be taken as evidence against OneToOne+. But subtleties in the data reveal that participants did *not* behave as if they were computing the more demanding ("significantly more") meaning. Rather, they behaved as if one extra yellow dot makes 'Most of the dots are yellow' true.

First, when participants had an easy way of determining that the cardinality of the yellow dots was greater than the cardinality of the nonyellow dots, they maintained that 10 yellow dots and 9 nonyellow dots was as an instance of most dots being yellow. This was the case on the Column Pairs Sorted trials in Experiment 1, in which participants relied on the length of the sorted columns as a proxy for number (Figures 7 & 8). In 94% of the relevant trials, participants treated scenes with 10 yellow and 9 blue dots as scenes described by 'Most of the dots are yellow' (see the data point furthest to the left in Figures 7 & 8 for Column Pairs Sorted trials).

Second, on the other three trial types (Scattered Random, Scattered Pairs, and Column Pairs Mixed) performance accorded with a psychophysical model that predicts that a single extra yellow dot suffices for the truth of 'Most of the dots are yellow'. Graphically, this can be seen in the fitted curves depicted in Figure 8 , where the curves do not cross the x-axis (chance performance) until a Weber Ratio of 1. This model predicts that participants will *tend* to answer the test question affirmatively, although the tendency may be slight, for any detectable positively signed difference between the yellow and nonyellow dots up to a Weber Ratio of 1—at which point the number of dots in each set is the same. So the best fit model of participants' performance predicts that any situation with at least one more yellow than nonyellow dot is a situation in which 'Most of the dots are yellow' counts as true.

Moreover, if participants understood 'most' as implying *significantly more* (as opposed to *at least one more*), then their accuracy should have systematically deviated from the model as Weber Ratio decreased. As the cardinality of the yellow dots became closer to the cardinality of the nonyellow dots, any participant who understood 'most' in the more demanding way should have refrained from assenting to 'Most of the dots are yellow' in a way *not* predicted by the model based on the less demanding meaning. As the Weber Ratio approaches 1, the model becomes a less accurate representation of anyone who understands 'most' as implying *significantly more*. To check for such a deviation, we calculated participant means for percent correct for each ratio bin across the three trial types (Scattered Random, Scattered Pairs, and Column Pairs Mixed) and plotted in Figure 9 the signed deviations of these means from the psychophysics model. Differences between percent correct and the psychophysics model were centered on zero and varied randomly from +6% to –6% with no tendency for these deviations to increase as Weber Ratio moved closer to 1. This means that participants behaved in accord with the psychophysics model, according to which one extra yellow dot suffices (up to the stochastic limits of the ANS to detect this difference) for judging that most dots are yellow.

General Discussion

In this paper, we have used psychophysical methods to adjudicate between hypotheses about 'most' that are equivalent by standard semantic tests. The meaning of 'most' can be described in terms of a relation (GreaterThan) that holds between the cardinalities of two sets, or in terms of a correspondence relation (OneToOne+) that holds between the individual elements of those sets. Because these characterizations are mathematically equivalent, there cannot be any situation that distinguishes them. So to determine which corresponds to the mental representations of competent speakers of English, one must find evidence that distinguishes hypotheses that are truth conditionally equivalent. In our view, the processes involved in evaluating a sentence with 'most' provide such evidence.

Our experimental data reveals two important points. First, despite our attempts to make evaluation of 'Most dots are yellow' easy given a OnetoOne+ meaning, we found no evidence that English speakers invoke algorithms that take advantage of one-to-one pairings of individuals in deciding whether a sentence using 'most' is true in a given situation. Second, and perhaps more positively, our data show that the Approximate Number System (ANS) is implicated in the algorithms used in computing the applicability of 'most'.

As noted above, this second finding is compatible with either of two scenarios. First, it is possible that adults represent the meaning of 'most' in terms of a comparison of magnitudes. Instead of (5a), perhaps the relevant meaning specification is (5a');

(5a) $\#\{x: Dot(x) \ \&\ Yellow(x)\} > \#\{x: Dot(x) \ \&\ \neg Yellow(x)\}$

(5a') $G\{x: Dot(x) \ \&\ Yellow(x)\} \ \square\ G\{x: Dot(x) \ \&\ \neg Yellow(x)\}$

where 'G' signifies a mapping (not from sets to cardinal numbers, but rather) from sets to Gaussian curves, each of which has a mean and a standard deviation, and '□' signifies a comparison (not of cardinal numbers, but rather) of Gaussians.

A second possibility is that (5a), interpreted standardly in terms of cardinal numbers, correctly represents the meaning of 'most'—but estimates of relevant cardinalities are supplied on-line by the ANS. Further experiments will be required to determine which of these two possibilities obtains (Halberda, Hunter, Lidz & Pietroski, in prep.).

A related question is whether participants in our experiment were actually computing the meaning of the stimulus question ('Are most of the dots yellow?') on every trial, or whether they might have converted it into a question with the word 'more'. Since each trial contained dots of exactly two colors, most of the dots were yellow iff there were more yellow dots than dots of the other color (blue). This is a serious issue. But note that our questions about 'most', and the related number-relevant representations, also apply to 'more'. Our results stand as an important test of these representations, independent of whether participants interpreted the task as asking a 'most'- question or a 'more'-question. And in one sense, 'most' must be deeply related to 'more': most Δs are Ψs iff the Δs that are Ψs *are more than* the Δs that are not Ψs. But how do speakers understand the claim that there are more of *these* than *those*?

Does it mean that these have a greater cardinality than those, or that these correspond OneToOne+ with those, or something else ? To the best of our knowledge, these issues remain unsettled by the literature on 'more'. That said, our participants did appear to engage the task as asking a 'most'-question, and they did not report translating this into a 'more'-question. In ongoing work, we examine whether performance (on 'Are most of the dots yellow') changes as a function of increasing the diversity of items in the contrast set—e.g. by presenting arrays with yellow, blue, green and red dots—since ‘more’-judgments and ‘most’-judgments can be teased apart in scenes with dots of several colors (Lidz, Halberda, Pietroski & Hunter, in prep.)

Finally, we want to stress that this kind of research, relating the cognitive science of number to the lexical semantics of natural language quantifiers, lets one ask questions about meaning that often go unaddressed for lack of relevant evidence. Every semanticist knows that for any given expression, there will be many truth-conditionally equivalent ways of describing its meaning. Given such equivalences, choosing among alternatives requires appeal to considerations, often concerning compositionality or theoretical parsimony. In the current case, we have argued that the evaluation procedures involved in understanding may provide some insight into the semantic representations themselves. And as we have argued, aspects of cognition that provide content for the linguistic system—in this case, the Approximate Number System—may place constraints on the representational vocabulary of the lexicon itself.

References

Barth, H. et al. (2003) The construction of large number representations in adults. *Cognition, 86*, 201 – 221.

Boolos, G. (1998): *Logic, Logic, and Logic*. Cambridge, MA: Harvard University Press.

Chierchia, G. & McConnell-Ginet, S. (2000). *Meaning and Grammar* (second edition). Cambridge, MA: MIT Press.

Cordes, S., Gelman, R., & Gallistel, C.R. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin and Review*, 8, 698-707.

Coren, S., Ward, L. M. & Enns, J. T. (1994). *Sensation and Perception*, 4th ed. Fort Worth: Harcourt Brace.

Davies, M. (1987). Tacit knowledge and semantic theory: Can a five per cent difference matter? *Mind* 96: 441–62.

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

Demopolous, W. (ed.) (1994): *Frege's Philosophy of Mathematics*. Cambridge, MA: Harvard.

Evans, G. 1981: Semantic theory and tacit knowledge. In S. Holtzman and C. Leich (eds), *Wittgenstein: To Follow a Rule* (London: Routledge and Kegan Paul).

Feigenson, L., Dehaene, S., & Spelke, E.S. (2004). Core systems of number. *Trends in Cognitive Sciences,* ***8***, 7, 307-314.

Fodor, J. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford: OUP.

Frege, G. (1879): *Begriffsschrift*. Halle: Louis Nebert. English translation in J.van Heijenoort (ed.), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*

(Cambridge, MA: Harvard University Press, 1967).

—(1884): Die Grundlagen der Arithmetik. Breslau: Wilhelm Koebner. English translation in J. L. Austin (trans.), *The Foundations of Arithmetic* (Oxford: Basil Blackwell, 1974).

—(1893, 1903): Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet, 2 vols. Jena: Pohle. English translation in M. Furth (trans.),

The Basic Laws of Arithmetic (Berkeley: University of California, 1967).

Gallistel, C.R. & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences*, 4(2), 59-65.

Hackl, M. (forthcoming). On the Grammar and Processing of Proportional Quantifiers: Most versus More Than Half. *Natural Language Semantics.*

Halberda, J. & Feigenson, L. (2007). Developmental change in the acuity of the "Number Sense": The approximate number system in 3-, 4-, 5-, 6-year-olds and adults. *Manuscript under review.*

Halberda, J., Sires, S.F. & Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science, 17*, 572-576.

Halberda, J., Taing, L. & Lidz, J. (2007). The age of 'most' comprehension and its potential dependence on counting ability in preschoolers. *Article under review.*

Higginbotham, J. (1985). On semantics. *Linguistic Inquiry* 16: 547-93.

Horty, J. (forthcoming). Frege on definitions: A case study of semantic content. Oxford: OUP.

Larson, R. & Segal, G. (1995). *Knowledge of Meaning*. Cambridge, MA: MIT Press.

Mandler, G. & Shebo, B.J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General,* 29, 224-236.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. WH Freeman and Company, New York NY.

Meck, W.H. & Church, R.M. (1983) A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, 9, 320-334.

Peacocke, C. (1986). Explanation in computational psychology: Language, perception and level 1.5. *Mind and Language* 1:101-123.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, 306, 499-503.

Pietroski, P. (2006). Induction and Comparison. *University of Maryland Working Papers in Linguistics* 15:157-190.

van Oeffelen, M.P. and Vos, P.G. (1982). A probabilistic model for the discrimination of visual number. *Perception and Psychophysics, 32*, 163 – 170.

Whalen, J., Gallistel, C.R. and Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science,* **10***(2)*, 130-137.

Wright, C. (1983): *Frege's Conception of Numbers as Objects*. Scots Philosophical Monographs, vol 2. Aberdeen: Aberdeen University Press.

Xu, F. & Spelke, E. (2000). Large number discrimination in 6-month-old infants. *Cogntion, 74*, B1-B11.

Acknowledgements

J.H. and J.L. devised the task; P.P, J.L., J.H., and T.H. defined the trial types of interest; T.H. implemented and ran the experiment; J.H. analyzed the data; P.P. and J.H. wrote the manuscript with input from J.L. and T.H. We thank Joanna Kochaniak for help in programming earlier versions of these studies.

Corresponding Author

Correspondence and requests for materials should be addressed to P.P. (pietro@umd.edu).

Rights of participants

Guidelines for testing human research subjects were followed as certified by the Johns Hopkins University and The University of Maryland Institutional Review Boards. Participants' rights were protected throughout.

Table 1. Pairwise comparison of trial types

Trial Types	F	*p*
Scattered Random-Scattered Pairs	.216	.651
Scattered Random-Column Pairs Mixed	.446	.518
Scattered Pairs-Column Pairs Mixed	.127	.728
Column Pairs Sorted- Scattered Random	152.17	.0001
Column Pairs Sorted- Scattered Pairs	193.89	.0001
Column Pairs Sorted- Column Pairs Mixed	131.66	.0001

Table 2. Parameter estimates from psychophysical model

Trial Type	R^2	Critical Weber Fraction	Nearest Whole-Number Ratio
Scattered Random	.9677	.32	3:4
Scattered Pairs	.8642	.33	3:4
Column Pairs Mixed	.9364	.30	3:4
Column Pairs Sorted	.9806	.04	25:26

Figure Legends

Figure 1. A scene of blue and yellow dots

Figure 2. Three scenes where the set of yellow dots and the set of blue dots are equinumerous, arranged as (a) column pairs sorted, (b) column pairs mixed, and (c) column non-paired

Figure 3. Two scenes where the yellow dots outnumber the blue dots arranged as (a) column pairs sorted, and (b) column pairs mixed.

Figure 4. Gaussian representations of the mental number line with linearly increasing means and standard deviations.

Figure 5. A tree diagramming the possible algorithms that could underlie the concept for 'most'.

Figure 6. Example displays from Experiment 1; (a) a Scattered Random trial, (b) a Scattered Pairs trial, (c) a Column Pairs Mixed trial, and (d) a Column Pairs Sorted trial.

Figure 7. Percent correct is displayed for each ratio as a function of trial type with area-controlled and size-controlled trials combined (mean of subject means ± SE)

Figure 8. Psychophysical model fits for each trial type.

Figure 9. The signed deviations of participants' responses on the *most*-task from the psychophysics model. Points designate the average deviation across Scattered Random, Scattered Pairs, and Column Pairs Mixed trials. Error bars designate the Standard Error of these deviations.

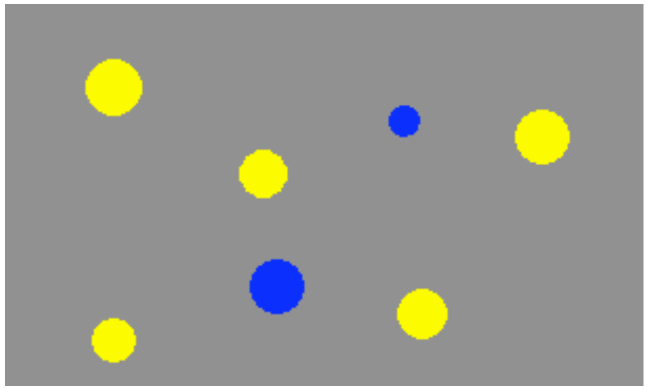


Figure 1

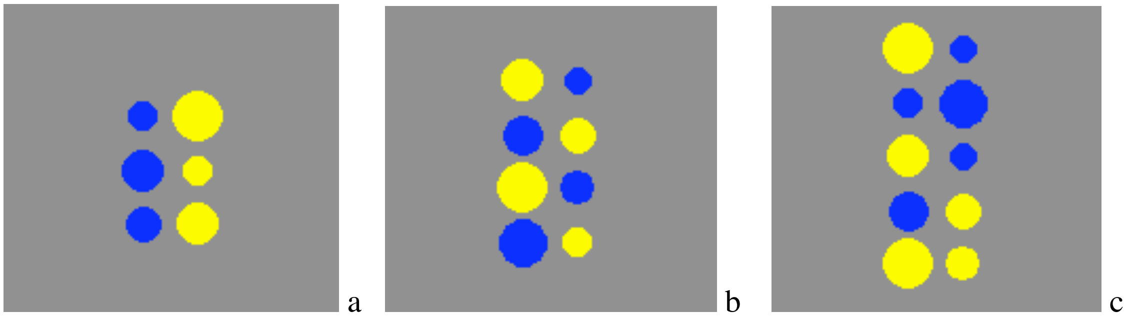


Figure 2

 a

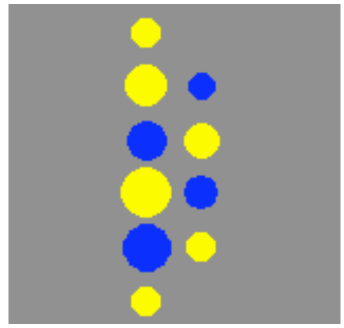
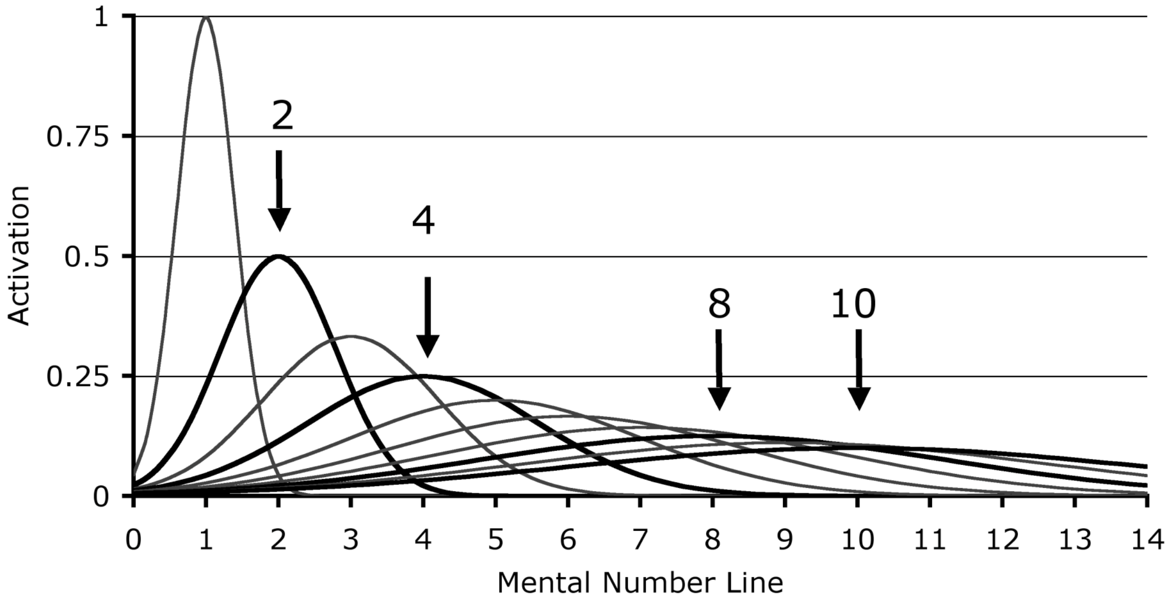
 b

Figure 3



1
0.75
0.5
0.25
0
Activation
2
4
8
10
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
Mental Number Line

Figure 4

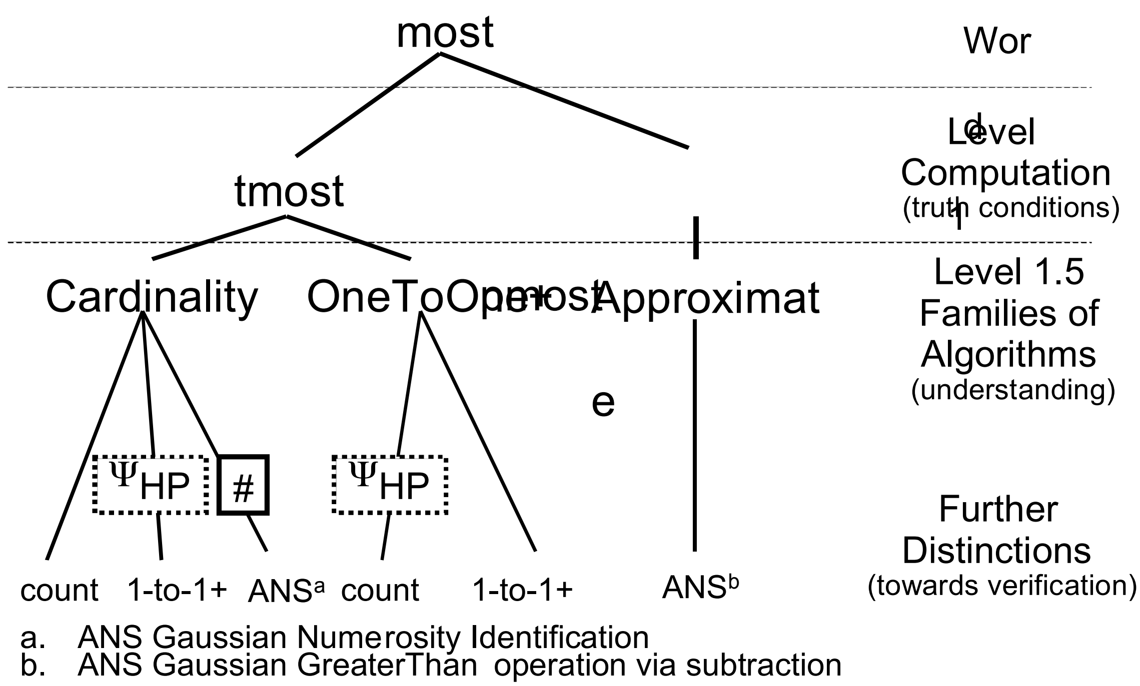

most
Wor
d
Level
Computation
(truth conditions)
tmost
Level 1.5
Families of
Algorithms
(understanding)
Cardinality
OneToOne+
most
Approximat
e
ΨHP
#
ΨHP
Further
Distinctions
(towards verification)
count
1-to-1+
ANSa
count
1-to-1+
ANSb
a. ANS Gaussian Numerosity Identification
b. ANS Gaussian GreaterThan operation via subtraction

Figure 5

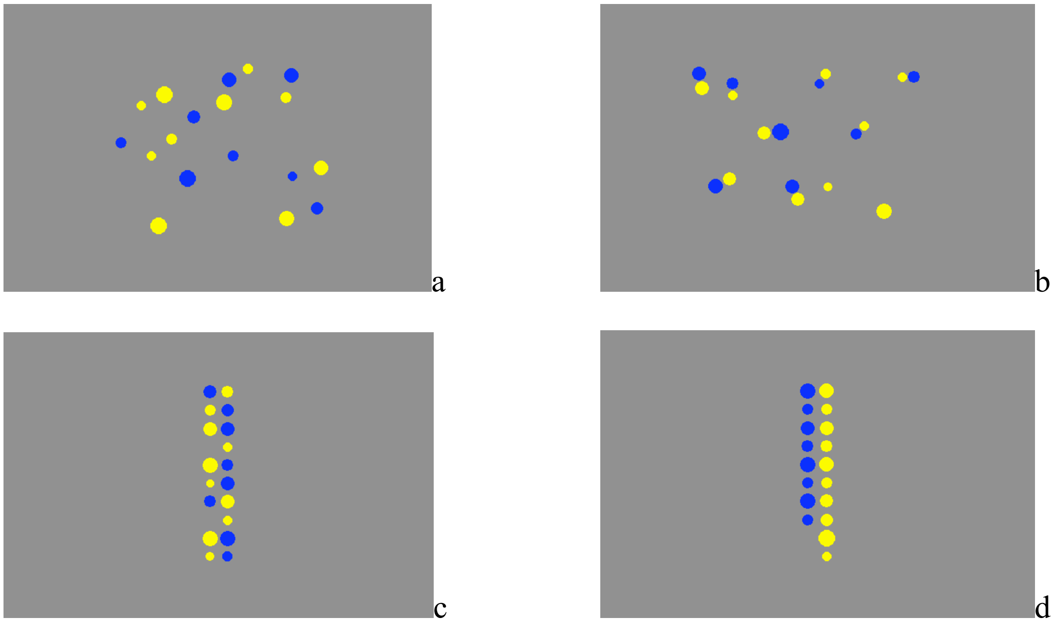


Figure 6

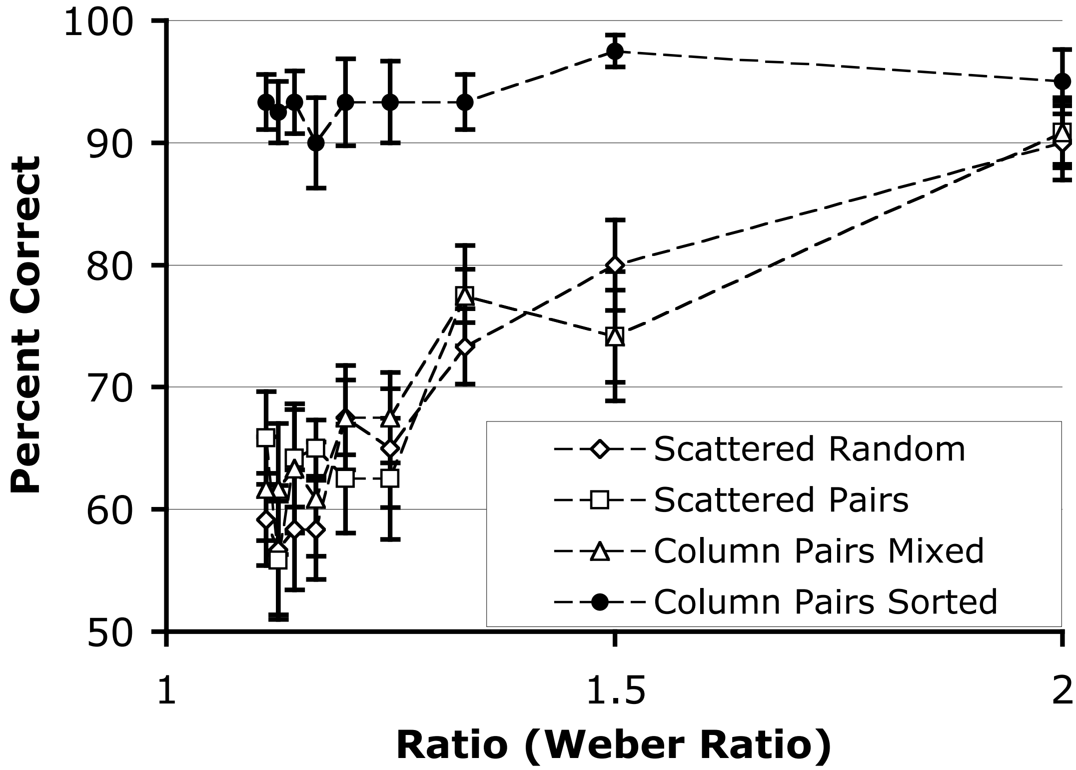
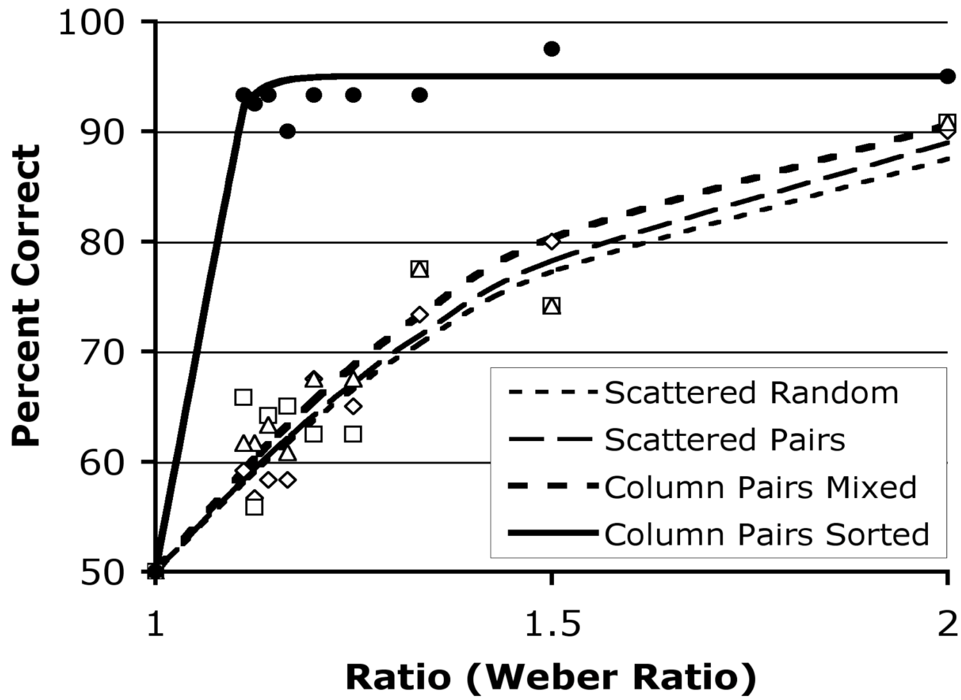

100
90
80
70
60
50
Percent Correct
1
1.5
2
Ratio (Weber Ratio)
Scattered Random
Scattered Pairs
Column Pairs Mixed
Column Pairs Sorted

Figure 7



100
90
80
70
60
50
Percent Correct
1
1.5
2
Ratio (Weber Ratio)
Scattered Random
Scattered Pairs
Column Pairs Mixed
Column Pairs Sorted

Figure 8

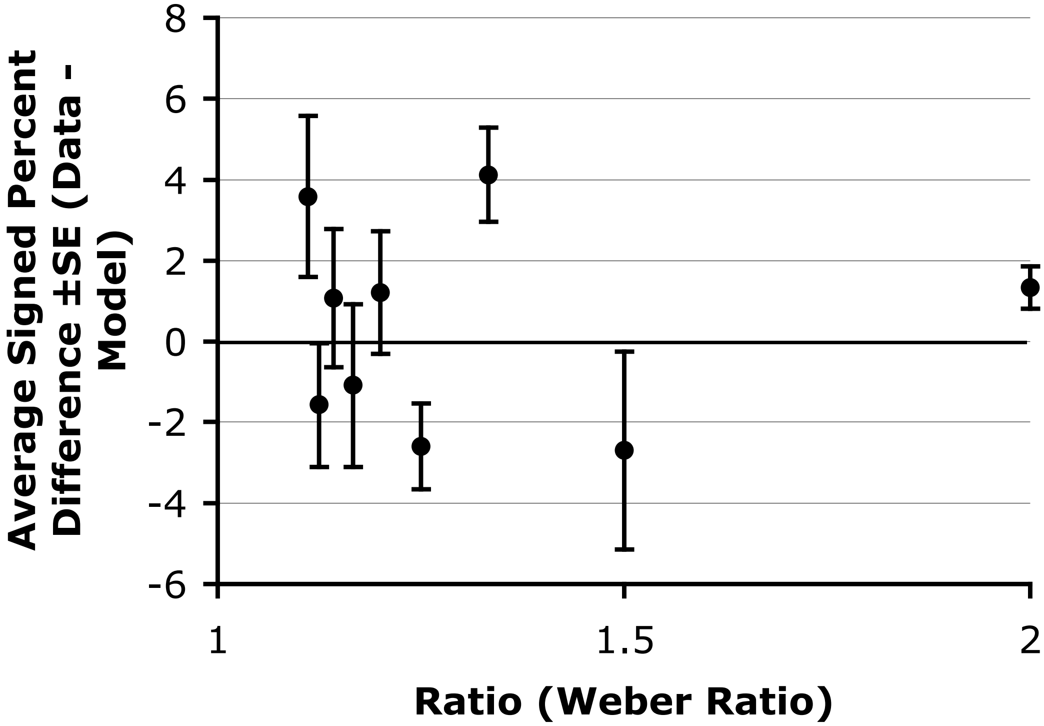

Average Signed Percent Difference ±SE (Data - Model)
8
6
4
2
0
-2
-4
-6
1
1.5
2
Ratio (Weber Ratio)

Figure 9

Notes

[i] Let us flag three potential complications. First, in the absence of salient dots, examples like 'Most dots are yellow' are naturally heard as generic—implying, roughly, that dots tend to be yellow. But in the experiments described below, the context of evaluation for 'Most of the dots are yellow' is transparent. Second, in ordinary discourse, instances of 'Most (of the) Δs are Ψ' often suggest that significantly more than half of the Δs are Ψ. This fact can be encoded as follows: $\#\{x:\Delta(x)\ \&\ \Psi(x)\} - \#\{x:\Delta(x)\ \&\ \neg\Psi(x)\} = n$; where n is a contextually determined number, capturing the idea that the number of Δs is *significantly* greater than the number of Ψs. But as noted in section four, competent speakers can and do judge that most of the dots are yellow in cases with nineteen dots, ten of which are yellow, suggesting that n=1. Moreover, statistical analysis reveals that participants in our experiments never understood 'Most of the dots are yellow' as implying that the yellow dots outnumbered the nonyellow dots by more than one. Third, we focus on uses of 'most' with pluralizable count nouns like 'dot', as opposed to mass nouns like 'water' and 'sand', as in as in 'Most of the sand is wet'. An adequate theory of the former should be extendable to the latter; and this might suggest that 'most' is understood in terms of 'more', or some notion of comparison not restricted to cardinalities. We return to this last point.

[ii] Correlatively, 'dots' is said to be the determiner's *internal* argument, while 'yellow' is said to be the *external* argument. This captures an analogy with sentences involving transitive verbs and names: 'Fido chased Garfield' has a constituency structure, [Fido [chased Garfield]], in which the verb takes 'Fido' as its *external* argument and 'Garfield' as its *internal* argument. Standard semantic theories reflect this linguistically important analogy.

[iii] As discussed below, one needn't worry about this if the aim is *simply* to provide a theory that associates each declarative sentence of English with a correct truth condition. In that case, proposed specifications are equivalent if and only if they are truth conditionally equivalent. But given other aims, other measures of equivalence are in order.

[iv] For discussion, see Wright (1983), Boolos (1998), and the essays in Demopolous (1994). At least prima facie, the left side of (HP) is an identity claim that logically implies the existence of at least one cardinal number, while the right side is a correspondence claim (concerning the elements of the sets in question) that does not logically imply the existence of any number. So while (HP) may be obvious, in some sense, it seems not to be a truth of *logic*.

[v] The background assumption, to which we return, is that understanding a 'most'-sentence involves constructing a complex mental representation of a truth condition; where this representation "embodies" a procedure for determining whether it is true or false. (See Horty [forthcoming] for discussion of the general idea, independent of verificationism, that a sentence S can embody a procedure intimately related to the meaning of S.) In many circumstances, the "direct" evaluation procedure embodied by a sentence will not be usable, thus prompting use of an alternative—like asking one's spouse. If an "indirect" procedure is reliable, one can use it to determine which answer the direct procedure determines. (There may also be cases where a direct procedure is us*able*, but all things considered, one is better off using a reliable alternative; if counting is possible, but you're apt to miscount, it may be better to ask your spouse.) But other things equal, if conditions for using a direct procedure of evaluation are favorable, then one presumably uses it. So if a procedure P is not used, in conditions favourable for its use in evaluating a 'most'-sentence, then it is implausible that understanding such a sentence leads to construction of a mental representation that embodies P.

[vi] Hackl (forthcoming) argues, in a similar spirit, that 'most' is not synonymous with 'more than half'; although his discussion is directed against a certain kind of lexical *decomposition*, in the service of suggesting that 'most' is better characterized as a superlative version of 'many'. By contrast, our remarks are compatible with the view that 'most' is a grammatically unstructured lexical atom (cp. Fodor [1998]), or a superlative version of 'more' (see section five). More generally, our inclination to construe semantic formalism psychologically is in the tradition of Evans (1981) and Davies (1987); see Larson and Segal (1995) for discussion. Peacocke (1986) talks explicitly in terms of algorithms; see note eight. And perhaps most pertinently for linguists, Chomsky's I-Language/E-Language constrast is based on the distinction between functions *in intension* and mere input-output pairs: I-languages are biologically instantiated *procedures*, not behavioral profiles; though in principle, distinct I-languages might associate a given range of sentential phrase markers with the same truth conditions.

[vii] For an exception, see Larson and Segal (1995). Though of course, many would agree that semantics should be viewed as a branch of human psychology; see, e.g., Higginbotham (1985), Chierchia and McConell-Ginet (2000).

[viii] Peacocke (1986) introduces the "Level 1.5" terminology to talk about algorithms that correspond to a certain kind of "information flow;" see also Davies (1987) on the "mirror constraint." But for our purposes, it is enough to contrast algorithms that require (representation and) comparison of cardinalities with algorithms that do not.

[ix] In principle, one could determine whether one number is bigger than another by counting out some things with each number, and seeing whether some of the former correspond one-to-one with all of the latter. But this would spoil the *point* of determining cardinalities as part of the process of determining whether most Δs are Ψ. One might also come to associate a general spatial arrangement with a limited range of cardinalities (e.g. three items typically form some kind of triangle, two items a line) and thereby avoid counting as a method for assessing cardinalities (Mandler & Shebo, 1982).

[x] The logic is second-order, permitting quantification into positions occupiable by predicates. But this cannot be foreign to a speaker who understand 'most'; see note 2. Of course, *deriving* the (Dedekind-Peano) axioms of arithmetic from (HP) requires definitions for arithmetic terms and appeal to a first number. But Frege defined zero as the number of things satisfying a logically contradictory property, and then given (HP), proved the following: zero has a unique successor, which has a unique successor, and so on; and these "descendants" of zero support proofs by (mathematical) induction. See Pietroski (2006) for discussion of the potential relevance for linguistics.

[xi] Given a meaningful claim, there can be many ways to find out if the claim is true; see Fodor (1998) for extended discussion. Even if 'Tim is a bachelor' *means* that Tim is an unmarried (marriageable) male, one can *find out* if Tim is a bachelor without checking marriage records or DNA. Though as this example illustrates, meaning can make a potential method of evaluation salient, especially with regard to *logical* vocabulary. To evaluate 'Tim is clever, and Tim is Australian', one might well follow the procedure suggested by the familiar truth table for conjunction. One might also ask Tim's friends. But given an understanding of 'and', one knows that a certain kind of evaluation procedure is available, and in some sense the "default" procedure to be used in the absence of shortcuts (i.e. "indirect" methods of evaluation, as discussed in note five). Likewise, we suggest: *if* 'most' is understood in terms of OneToOne+, then competent speakers should at least be inclined to use a correspondence evaluation procedure.

[xii] We take no stand on whether young children or nonhuman animals can have such representations. And in adult humans, such representations presumably exhibit some kind of recursive structure. Absent special training, thinkers may not have *atomic* representations of ninety—or even nine, which may be represented as the successor of eight. (Though we may well have an atomic representation of three.) Or if integer words are understood as second-order

predicates, as opposed to designators, 'nine' may be understood as a predicate that applies to some things, the Xs, iff there are some things, the Ys, such that: there are eight Ys; every Y is an X; and exactly one of the Xs is not a Y.

[xiii] Even the *initial* representations/curves of the ANS—those with the smallest standard deviations, indicated towards the left in figure 4—are not representations of precise cardinalities; although only rarely will the ANS fail to distinguish a scene with (exactly) two perceptible items from an otherwise similar scene with (exactly) one perceptible item. The mental number line might also be modelled as logarithmically organized with constant standard deviation which would change the look of the Gaussian curves in Figure 4, but they would remain Gaussian. Either format results in the hallmark property of the ANS: discrimination of two quantities is a function of their ratio (Weber's Law). Here we will assume the linear format, as it has traditionally been the more dominant model (e.g., Cordes et al., 2001; Gallistel & Gelman, 2000; Meck & Church, 1983, Whalen et al., 1999).

[xiv] Recall that 'most' is essentially comparative—most Δs are Ψ iff $\#\{x: \Delta(x) \& \Psi(x)\} > \#\{x: \Delta(x) \& \neg\Psi(x)\}$—in contrast with (firstorderizable) determiners like 'every'.

[xv] Consider a scene with exactly 100 dots, 51 of which are yellow. Any such scene satisfies the condition imposed by 'Most dots are yellow', given how numerate adults understand 'most'. And this is so, whether or not there is adequate opportunity to count the dots. We offer independent evidence, below, for this claim about adults. But it seems clear that if 51% of the trees in a forest fall with noone to count them, it is still *true* that most of the trees fell.

[xvi] The latter seems especially implausible, and our experiment tells against the former. But this is not to say that such procedures are impossible for competent speakers (see note eleven).

[xvii] Because comparisons made in the ANS are ratio-dependent, the discriminability of 6 yellow versus 8 blue dots would match that for 12 yellow versus 16 blue dots. If subjects are relying on the ANS to evaluate 'most' (for either *tmost* or *pmost*) then there should be no effect of the total number of items involved in a display and only an effect of the ratio between the two sets. Two possible reasons for the slight decrease in performance (i.e. a decrease of –.365% per dot) as a function of increasing Total number of dots in the display are that increased dots led to increased perceptual crowding making it harder for the ANS to generate accurate estimates of numerosity or that increasing Total number of dots in the display led to an increased tendency for participants to randomly guess, as if participants felt that when there were many dots in the display the task was harder (a similar effect has been seen in other tasks that engage the ANS for purposes of speeded comparative judgements: Pica et al, 2004). Thus, the present analysis

shows no evidence for an error prone OneToOne+ algorithm and remains consistent with the hypothesis that participants relied on the stochastic representations of the ANS.