The Meaning of 'Most': semantics, numerosity, and psychology

Paul Pietroski, *University of Maryland*

Jeffrey Lidz, *University of Maryland*

Tim Hunter, *University of Maryland*

Justin Halberda, *Johns Hopkins University*

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

Abstract

To a first approximation, 'Most of the dots are yellow' means that the number of yellow dots exceeds the number of nonyellow dots. A less numeristic paraphrase is that some of the yellow dots correspond one-to-one with all of the nonyellow dots. Indeed, given current semantic theories, there are many ways to describe the meaning of 'most' (see Hackl [2008]). We offer a framework for distinguishing such descriptions, interpreted as psychological hypotheses. We then present data that tells against the theoretically attractive 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 yellow dots and blue dots were shown briefly on a computer screen. Displays were manipulated to vary the ease of using a "one-to-one with remainder" strategy, for evaluating the sentence, and the ease of using the Approximate Number System (ANS) as part of a strategy involving comparisons of (approximations of) cardinalities. Results suggest that the target sentence is understood in terms of cardinality comparison, even when counting was not possible; although the use of such data requires care, in thinking about how meaning is related to verification. We discuss 'most' 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 speakers deploy in comprehension.

The Meaning of 'Most': semantics, numerosity, and psychology

How is the word 'most' related to human capacities for detecting and comparing numerosities? One might think the answer is obvious, and explicit in standard semantic theories: 'most' is understood in terms of a capacity to compare cardinal numbers; e.g., 'Most of the dots are yellow' means that the *number* of yellow dots is *greater* than the *number* of nonyellow dots. But there are other possibilities for how competent speakers understand 'most', and we offer experimental evidence that tells against some initially attractive hypotheses.[1] By focusing on one lexical item in this way, we hope 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 begin by characterizing the contribution of 'most' to *truth conditions* of sentences of the form 'Most (of the) Δs are Ψ'. At this level of analysis, there are many equivalent characterizations, as described below. Our aim is to give some of these formal distinctions empirical bite, in a way that permits adjudication of distinct hypotheses about how speakers *understand* 'most'. We want to know how speakers *represent* the truth-conditional contribution of this word. But the initial focus on what is represented provides needed background for our questions concerning the representational format linked to 'most'.

---

[1] Hackl (2008) raises important questions concerning how cardinality comparisons are represented, and how this relates to issues concerning natural language syntax (and potential lexical decomposition), in the context of experimental evidence concerning contrasts between 'most' and 'more than half'. Our questions here are in the same spirit. But we are less concerned with syntax, and we do not assume that 'most' is understood (one way or another) in terms of cardinality comparison. Our investigation begins with the question of how to justify this assumption empirically, with the aim of setting out a broader framework for posing such questions in testable ways. But we view this as a very friendly extension of Hackl's approach.

The rest of this introduction outlines our discussion. Section one contrasts two *ways of specifying* a truth condition for 'Most of the dots are yellow': comparing numbers (of yellow dots and nonyellow dots), *vs.* comparing yellow and nonyellow dots for a certain kind of correspondence relation. Sections two and three are about how such specifications are related to understanding, verification, and distinctions between algorithms and functions computed; cp. Dummett (1973), Marr (1982), Peacocke (1986, 1992), Chomsky (1986), Horty (2007). The idea is that 'most' is semantically linked to a *procedure* that determines answers to yes/no questions of the form 'Most Δs are Ψ?', given subprocedures that determine whether something is a Δ and whether something is a Ψ. From this perspective, to understand 'most' is to recognize it as an instruction to generate (a representation of) a certain *algorithm* with binary (yes/no) outputs. A thinker may often be unable to execute this algorithm and use it as a *verification strategy*; and there may often be better ways of answering the question at hand. But in suitably controlled circumstances, we claim, the 'most'-algorithm will be used as a default verification strategy.

Sections four and five review independent evidence that humans have the cognitive resources to implement *both* cardinality-comparison algorithms (via approximate number representations) *and* correspondence algorithms for 'most'. Section six describes an experiment in which displays of dots varied across trials, in ways that made it easier or harder to employ a correspondence algorithm as a verification strategy. There was no evidence of subjects using such a strategy. By contrast, there was clear evidence of subjects of comparing cardinalities, via the Approximate Number System (ANS). In sections seven and eight, we discuss some potential concerns about the task. But detailed analysis confirms our conclusion that even when speakers are forced to estimate the relevant cardinalities, they understand 'most' in terms of whether one number is greater than another.

1. Background Semantics

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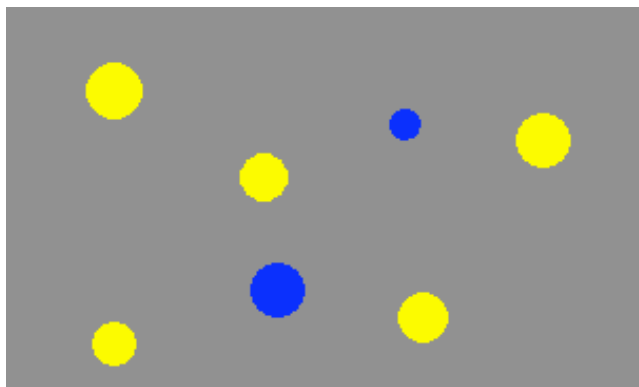
Figure 1

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 figure 1, $\#\{x: \text{Dot}(x) \,\&\, \text{Yellow}(x)\} = 5$, while $\#\{x: \text{Dot}(x) \,\&\, \neg\text{Yellow}(x)\} = 2$. So perhaps 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.[2]

---

[2] Let us flag three complications. First, absent salient dots, 'Most dots are yellow' is heard as generic—implying that dots tend to be yellow. But in our experiment, the relevant dots are obvious, and the context is clearly nongeneric. Second, instances of 'Most (of the) Δs are Ψ' often suggest that significantly more than half of the Δs are Ψ. But as discussed below, our experiment provides independent reason for treating this as a pragmatic effect. Third, we focus on pluralizable count nouns like 'dot'. But since an adequate theory must be extendable to mass nouns, as in 'Most of the sand is wet', one might think that 'most' is understood in terms of a comparitve notion whose meaning is *not* specified in terms of cardinalities. We return to this point in our final section and in other work.

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).[3]

(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)} ⊆ {x: Yellow(x)}. Likewise, one can say that 'some' signifies nonempty intersection, and hence that (3) is true iff {x: Dot(x)} {x: Yellow(x)} ≠ Ø. But one can also describe the semantic contributions of 'every' and 'some' in terms of cardinalities: every Δ is Ψ iff #{x: Δ(x) & ¬Ψ(x)} = 0; some Δ is Ψ iff #{x: Δ(x) & Ψ(x)} > 0. From this perspective, 'five' is a special case of 'some', since (exactly) five Δs are Ψ iff #{x: Δ(x) & Ψ(x)} = 5. This invites the characterization of 'most' noted above: most Δs are Ψ iff #{x: Δ(x) & Ψ(x)} > #{x: Δ(x) & ¬Ψ(x)}. But while these biconditionals may correctly and usefully describe the contributions of determiners to sentential truth conditions, psychological questions remain. In particular, what representations do competent speakers generate in understanding determiners?

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. It seems like overintellectualization to say that sentence (2) *means* that zero is the number of dots that are not yellow, or that (3) *means* that the set of yellow dots has a cardinality greater than zero. Indeed, one might describe the

[3] The noun 'dots' is the determiner's *internal* argument, while 'are yellow' is the *external* argument—on the model of [Fido [chased Garfield]], in which 'Garfield' is the *internal* and 'Fido' as the *external* argument of the verb.

semantic roles of 'every' and 'some' in first-order terms, without reference to sets: every Δ is Ψ iff ∀x:Δx(Ψx), some Δ is Ψ iff ∃x:Δx(Ψx). Analogous issues arise with regard to 'most'.

Theorists *can* capture the truth-conditional contribution of 'most' without reference to numbers, by appealing instead to one-to-one correspondence. For the notions of cardinality and one-to-one correspondence are intimately related. The yellow dots and the blue dots have the *same* cardinality, as in each of the scenes in Figure 2, iff the yellow dots correspond one-to-one with the blue dots.
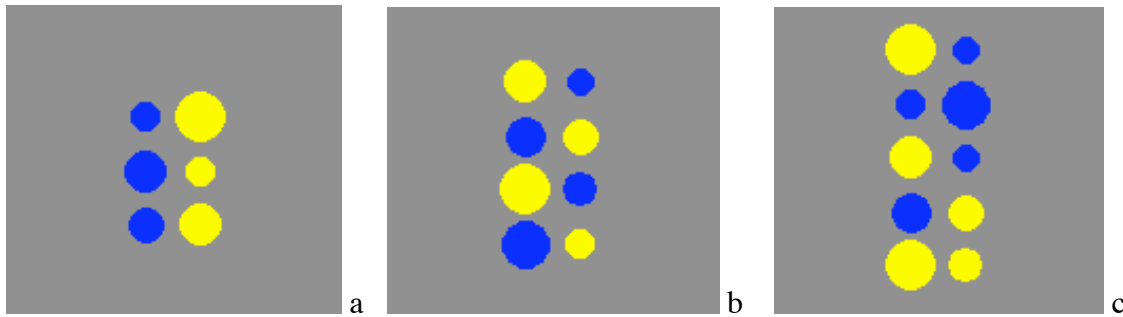

Figure 2

More generally, 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 return to the relevant generalization, often called "Hume's Principle," which lies at the heart of arithmetic.

(HP) $\#\{x: \Delta(x)\} = \#\{x: \Psi(x)\}$ iff 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.[4] 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 determine which of two collections has the *greater* cardinality—without being able to recognize, compare, and distinguish cardinalities. Counting is not required to determine if some things outnumber some other things. In each scene in Figure 3, there are more yellow dots than blue dots.
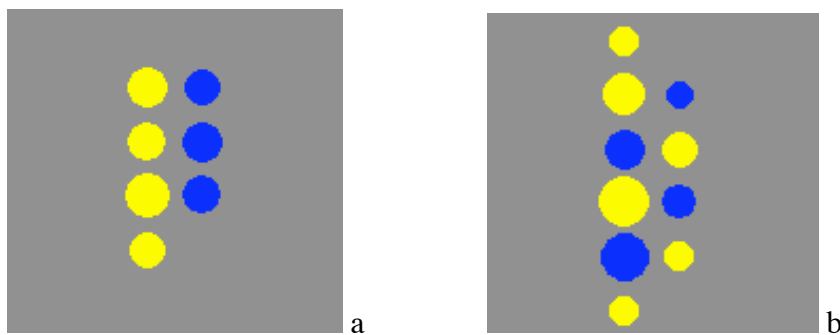

a   b

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 bears the

---

[4] 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*.

OneToOne relation to the set of blue dots. And theorists can define a second correspondence

relation, OneToOnePlus, as shown below.

OneToOnePlus[$\{x: \Delta(x)\}, \{x: \Psi(x)\}$] iff

(i) for some set **s**, **s** $\subset$ $\{x: \Delta(x)\}$ & OneToOne[**s**, $\{x: \Psi(x)\}$]

and (ii) ¬OneToOne[$\{x: \Delta(x)\}, \{x: \Psi(x)\}$]

For finite domains, clause (ii) is otiose: if a proper subset of $\{x: \Delta(x)\}$ bears the OneToOne

relation to $\{x: \Psi(x)\}$, then $\{x: \Delta(x)\}$ is not itself so related to $\{x: \Psi(x)\}$. But in any case, the

following generalization is correct.

GreaterThan[#$\{x: \Delta(x)\}$, #$\{x: \Psi(x)\}$] iff OneToOnePlus[$\{x: \Delta(x)\}, \{x: \Psi(x)\}$]

So (5) is true iff OneToOnePlus[$\{x: Dot(x)$ & $Yellow(x)\}, \{x: Dot(x)$ & ¬$Yellow(x)\}$].[5]

(5) Most dots are yellow

Assuming finitely many dots, this condition is met iff for some proper subset **s** of the yellow

dots, OneToOne[**s**, $\{x: Dot(x)$ & ¬$Yellow(x)\}$]

Perhaps, then, (5) is not *understood as* a numeristic claim according to which one

cardinality exceeds another. Though likewise, (5) may not be *understood as* a correspondence

claim. Given formally distinct but truth-conditionally equivalent ways of describing the semantic

contribution 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, much as there is no fact of

the matter about whether temperatures should be measured in Farenheit or Centigrade; cp. Quine

---

[5] For infinite domains, clause (ii) is needed to preserve the generalization. For example, while the even numbers

form a proper subset of the naturals, the naturals do not have a greater cardinality than the evens. But this Cantorian

fact supports the point that counting is *not* required to determine if some things outnumber some other things.

(1960), Davidson (1974). 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 invite, in some way that needs to be made more explicit, psychological hypotheses that differ in plausibility.

In particular, one might think it is *less* plausible that (5) is understood in terms of cardinalities. Competent speakers can reliably answer questions involving 'most' in situations that preclude counting. Moreover, in other work, we have found that some children who seem to lack exact cardinality concepts still seem to understand 'most' appropriately (Halberda, Taing and Lidz, 2007). Such facts are hardly decisive. Adults may have various methods for verifying cardinality claims, when counting is impossible or not worth the bother; and children may lack full competence. But distinctions among truth-conditionally equivalent specifications can at least suggest different *algorithms* for determining the truth/falsity of (5). And one can hypothesize that competent speakers associate sentences like (5) with algorithms of a certain sort.

2. Meaning and Verification

For these purposes, let's assume that *understanding* sentences of the form 'Most Δs are Ψ' requires (exercise of) a general capacity to represent the truth conditions of such sentences accurately and compositionally. The question here is whether such understanding requires something more, like a capacity to generate *canonical specifications of* the relevant truth conditions; where for any such sentence, a canonical specification corresponds to a certain *effective procedure* that yields conditional specifications of truth values, given decisions about how to classify things as Δs/not-Δs and Ψs/not-Ψs.

One could certainly invent a communication system in which the sound of (5)

(5) Most dots are yellow

indicates the truth condition that is indicated equally well by (5a) and (5b).

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

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

A competent user of the system might be free to represent this truth condition in any workable way, and express any such representation—regardless of format—with the sound of (5). But one could also invent a language in which this sound is, by stipulation, an instruction to generate a representation of the form shown in (5a). We think the natural phenomenon of understanding is psychologically demanding in this sense. But a canonical specification/procedure need not be used as a *verification strategy*; see, e.g., Dummett (1973), Horty (2007).

A canonical specification/procedure is a *way of computing* a truth condition. And one can make a yes/no judgment concerning the *truth condition computed*—in Fregean mode, one can make a judgment about *which* truth value is the one conditionally specified—without executing the mode of composition in question. Given (5a) as a way of computing the truth condition of (5), one can employ many "methods" (or verification strategies) to make the corresponding yes/no judgment. Depending on one's capacities and background beliefs/desires, one might: count and compare cardinalities; pair off dots and check for remainders; ask a friend, roll some dice, use one of these methods for a small sample of dots and extrapolate, or whatever.

Put another way, understanding (5) may be a matter of perceiving this sentence as one whose truth or falsity is specified in a certain way. But this is compatible with endlessly many strategies for judging whether the sentence is true. For even if one perceives (5) as posing the question indicated in (5a), as opposed to (5b), one might seek an answer in many ways.

The two procedures we focus on, cardinality-comparison and OneToOnePlus-assessment, can never disagree: there is no conceivable scenario in which these algorithms yield different results. Of course, actual attempts to execute the algorithms may fail in different ways in different circumstances. But taking the outputs to be conditional specifications of truth values, for any instance of 'Most Δs are Ψ', the two procedures cannot ever yield specifications that specify different truth values. Nonetheless, the "truth-procedures" differ. One can imagine creatures who cannot represent cardinalities, and so cannot associate (5) with the first procedure. Likewise, one can imagine creatures who lack the representational resources to associate (5) with the second procedure. One can also imagine creatures who have the cognitive resources required to associate (5) with *either* procedure, but in fact understand (5) in exactly one way, sometimes using the other procedure as a verification strategy.

Let us stress: our view is not that meaning is verification, or that studying verification strategies is a generally useful way to study meanings. We take it as given that whatever a declarative sentence means, there will be endlessly many ways of judging (in a given setting) whether or not the sentence is true. But we do find it plausible that at least with regard to "logical" vocabulary, meanings are individuated at least as finely as truth-procedures (see Frege [1892] and Church [1941] on sentential "functions-in-intension"), which can but need not be used as verification strategies.[6] One way of vindicating this old idea is to independently describe situations that lead competent speakers to *use* meaning-specifying effective procedures *as* default

---

[6] Horty (2007) offers recent and penetrating discussion, drawing on Dummett (1973) and others, explicitly in the context of algorithms/procedures and their relation to Frege's notion of definition. Cp. Chomsky's (1986) discussion of I-languages, with 'I' connoting intensional/procedural characterizations of functions. In this context, intensions should not be identified with sets of possible worlds, as Church's own (algorithm-focused) discussion makes clear.

verification strategies. If such situations can be empirically identified, this could also support (nondemonstrative but confirmable) inferences about meaning from facts about verification *in these special situations*.

Of course, there can be no guarantee that different speakers associate 'most' with the same canonical specification/procedure, or even that any one speaker does so across contexts. But one can try to argue empirically that certain algorithms, which initially seem like plausible candidates, are not used in situations that positively invite their use. For if certain algortithms are never used as default verification procedures, then at least to this extent, competent speakers do not vary. More importantly, experimental evidence may reveal the use of verification strategies that are *at odds with* certain hypotheses about canonical specifications/procedures, given independently confirmed models of the cognitive systems that support the strategies.

To foreshadow, imagine discovering that adults often rely on the Approximate Number System (ANS) in judging whether or not most of the dots are yellow. Independently attractive models of this system suggest that it has no representations for single individuals; rather, the ANS generates estimates of the cardinalities exhibited by one-or-more things. In which case, the meaning of 'most' must permit verification via a cardinality-estimating system that does not represent single individuals. This turns out to be a substantive constraint, as we'll argue in section five, since understanding 'most' in terms of correspondence would require a mechanism for converting representations like (5b) into representations like (5a).

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

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

3. Analogy to Marr

The analogy to Marr's (1982) contrast, between Level One (computational) and Level Two (algorithmic) questions, is intended. Given a system that seems to be performing computations of some kind, theorists 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.

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 Bob (who cannot count) has always defined 'most' in terms of one-to-one correspondence. By contrast, Chris has an ANS that 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 Bob; but in close cases, say 8 versus 7, Chris cannot tell the difference and so defers to others. Alex, Bob, and Chris can communicate. Indeed, they never disagree about whether sentences of the form 'Most Δs are Ψ' are true. In this sense, they understand each other, at least in a Level One way. But their psychologies differ in important respects.

Given some assumptions about the kinds of sentential properties that competent speakers can recognize—say, the truth conditions of declarative sentences—one can ask *how* speakers recognize these properties. In practice, semanticists usually 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.[7] For often, it is hard enough to find *one* good way of

---

[7] 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).

characterizing a "composition function" that speakers *somehow* compute. 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.

Similarly, one can and should distinguish the truth conditional contribution of a word like 'most' from how that contribution is represented by competent speakers. It is often useful to begin with a proposal about what the truth conditional contribution *is*; and for these initial purposes, truth conditionally equivalent representations are equivalent. The corresponding Level One task—of describing the 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 relevant syntactic constraints—is far from trivial. But depending on the *theorist's choice* of formal notation, it might be enough to say that each determiner indicates a certain function from pairs of predicates to truth conditions. In which case, endlessly many formally distinct lexical specifications will be acceptable. But if the task is to say how *speakers understand* words, with the aim of saying how linguistic comprehension frames the task of evaluating sentences, theorists must get beyond Level One questions about which function a given word indicates.

It may not be possible, at this early stage of inquiry, to formulate defensible Level Two 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'.[8] Do they represent cardinalities, and a relation between them, or not?. Of course, there is more than one cardinality-comparison

---

[8] 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.

algorithm; and there can be no guarantee that all speakers specify the meaning of 'most' in terms of the same one.[9] But theorists can distinguish two *classes* of algorithms for determining whether or not most dots are yellow: algorithms which require representations of two cardinalities and a subprocedure for determining whether the first is greater than the second; and algorithms which require dot-representations that support a procedure for determining suitable correspondences (without needing to represent cardinalities).

Given this distinction, one can ask for each class of algorithms—i.e., for each abstract "way" of computing the relevant function—if there is *some* psychologically plausible way of computing the function in that (abstract) way. In this context, let us return to an earlier point. One might think that cardinality-comparison specifications/procedures are psychologically *im*plausible. Independent evidence suggests that counting is often slower and more laborious than evaluating sentences with 'most'. And in situations where items are presented too briefly to count, yet the subject can still determine whether or not most Δs are Ψs, appealing to a OneToOnePlus algorithm might seem especially attractive.

More generally, one might think that humans who cannot count understand 'most' in terms of correspondence as opposed to cardinalities. If this is correct, parsimony would suggest that adults who can count still *understand* 'most' in the same way. Perhaps counting provides a

---

[9] The number of yellow dots might be determined by counting, or by recognizing a pattern already associated with a number. (One might 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; see Mandler & Shebo [1982], LeCorre & Carey [2007].) 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. And so on.

new way of determining whether some things (like yellow dots) bear the OneToOnePlus relation

to some other things (like nonyellow dots). This suggestion coheres with certain foundational

considerations that may well be germane to semantics and logic. Recall the generalization (HP),

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

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.[10] Given this logic, (HP) encapsulates

arithmetic. The intuitive simplicity of one-to-one correspondence, and its relation to the

foundations of arithmetic, thus adds to the attractions of saying that 'most' (along with other

counting/cardinality expressions) is understood in terms of one-to-one correspondence.

Given these considerations, one might expect speakers to jump at the chance of

evaluating 'Most dots are yellow' *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 OneToOnePlus strategy

would seem to be a plausible candidate for speakers who (for whatever reason) cannot count the

---

[10] 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.

relevant dots . Our experiment, described in section six, suggests that this strategy is not available to participants. But before turning to the details, we need to review some relevant literature concerning the psychology of number.

4. 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) OneToOnePlus[{x: Dot(x) & Yellow(x)}, {x: Dot(x) & ¬Yellow(x)}]

involve two kinds of representational resources for which there is independent evidence. In this sense, neither class of algorithms requires novel cognition. At least for adults who know the meanings of number words, there clearly are representations of exact cardinal values—like 'seven' and 'sixty'—that *could* support a cardinality specification/procedure for 'most'. And demonstrations of object-tracking in infants (Wynn, 1992; Feigenson, 2005) have revealed a system that can detect one-to-one correspondences, at least in certain situations. But we must also consider a third cognitive resource, the Approximate Number System, that can provide numerical content for certain representations. We wanted to discuss this system separately, with the distinction between (5a) and (5b) in place, without prejudging questions about whether and how speakers can use the ANS to verify 'most'-claims.

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 widely endorsed model of the representations generated by the ANS: 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.[11] We return to this point.
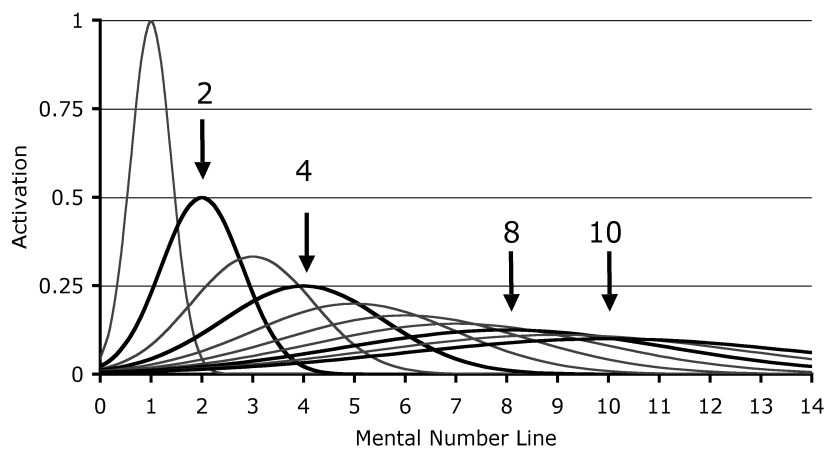

Figure 4

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

---

[11] 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 and either model would be applicable to the simple discrimination task we rely on (e.g., Cordes et al., 2001; Gallistel & Gelman, 2000; Meck & Church, 1983, Whalen et al., 1999).

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.[12] 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, 2007). 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 very quickly (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-

---

[12] 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'.

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 forming a smooth Gaussian curve centered on 'twenty-five'. (Halberda et al, 2006; Le Corre & Carey, 2007; 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.

Let us return, now, to the meaning of 'most'. As noted above, the ANS will not deliver a representation of something *as* exactly one. This system does not generate a representation of unity or any representation of a "minimal" difference between distinct cardinalities; cp. Leslie, Gallistel, and Gelman (2008). If only for this reason, speakers cannot rely on the representations of the ANS to implement a OneToOnePlus algorithm, for understanding or evaluation. Given one-or-more yellow dots, the ANS can represent "it-or-them" *as* oneish; and in principle, the yellow dot(s) so-represented might be paired with some similarly represented blue dot(s), until only unrepresented yellows or blues remain. Thus, one can imagine a "OneishToOneishPlus" algorithm. But as we'll see, the data suggest a *very* different use of the ANS.

Nonetheless, representations of the ANS do represent things, stochastically, as ordered in a way that seems numerical. And these representations can be mapped onto discrete number words. This suggests two possible ways of using ANS-representations: first, in a canonical specification/procedure of the truth conditions for 'most'-sentences, without any independent appeal to (concepts of) cardinal numbers; or second, in a frequently used verification strategy that naturally interfaces with a meaning canonically specified in terms of cardinality comparison. The first possibility deserves extended discussion. But it also raises technical difficulties that go

well beyond the scope of this paper.[13] So here, we focus on the second option. Correspondingly, our subjects are adult speakers of English who have presumably linked the Gaussian approximate number representations to the discrete number words.

As a way of fleshing out this second possibility, of speakers using their ANS in verification, one might imagine a "numeralizing waystation"—perhaps restricted to speakers like our subjects—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

---

[13] We would be led into discussions of partial functions, vagueness, supervaluation, uncertainities, and secondary qualities. A tempting idea is to introduce a metalanguage notion 'Exceeds(x, y)', whose variables range over cardinal numbers *and* ANS representation types; where unlike any pair of numbers, a pair of ANS representations that lie within a specified range of "unreliable discriminability" can be such that neither Exceeds the other. In ideal conditions, each ANS representation will have a Mode(/Mean/Median)-number: the likeliest (/average/middle) number of things to trigger the representation. But one can define indifference notions so that ANS-representations with different M-numbers are indifferent with regard to 'Exceeds(x, y)'. Theorists might then represent the meaning of the word used by numerate adults ('Most$_{num}$') as a precisification of the meaning of an earlier word, 'Most$_{innum}$', sucb that: when most of the dots are yellow, and it is not a close call, 'Most$_{innum}$ 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, 'Most$_{innum}$ of the dots are yellow' is false; and sometimes, as when 51% or 49% of the dots are yellow, 'most$_{innum}$ of the dots are yellow' is neither true nor false. But whatever the details, two basic questions will remain: does the metalanguage expression 'Exceeds(x, y)' indicate a total function from pairs number-like entities to truth values, so that the result is false when the values of 'x' and 'y' are indifferent ANS-representations; and how should theorists think about the relation of their invented metalanguage to the concepts of speakers? We cannot discuss such questions here, important though they are.

waystation that interfaces between the ANS and other cognitive systems, 'most'-sentences could be *understood* as in (5a) yet often verified (imperfectly) via the ANS.

5. Putting the Pieces Together

By way of summarizing to this point, let's stipulate that the 'most'-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: GreaterThan[#{x: $\Delta(x)$ & $\Psi(x)$}, #{x: $\Delta(x)$ & $\neg\Psi(x)$}]; OneToOnePlus[{x: $\Delta(x)$ & $\Psi(x)$}, {x: $\Delta(x)$ & $\neg\Psi(x)$}]; etc. From an extensional (E-language) standpoint, these different ways of notating the 'most'-function are just that: notational variants. But from an intensional (I-language) perspective, the different notational choices suggest various possibilities for canonical specifications/procedures in terms of which speakers understand 'most'. Some of these possibilities are depicted in Figure 5.

Speakers who associate (the sound of) 'most' with the 'most'-function might understand 'Most of the dots are yellow' as a claim that logically implies cardinalities, or as a less numerically loaded correspondence claim . 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. Positing such a waystation is indicated, in figure 5, with the box containing '#'. We also note, for completeness, two possibilities that seem especially unlikely: given a psychologically realized version of Hume's Principle, indicated with 'ΨHP', a thinker might *understand* 'most' in terms of cardinality but (whenever possible) *verify* by checking for OneToOnePlus correspondence—or conversely, understand 'most' in terms of OneToOnePlus but default to verifying by counting.[14]

---

[14] 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.
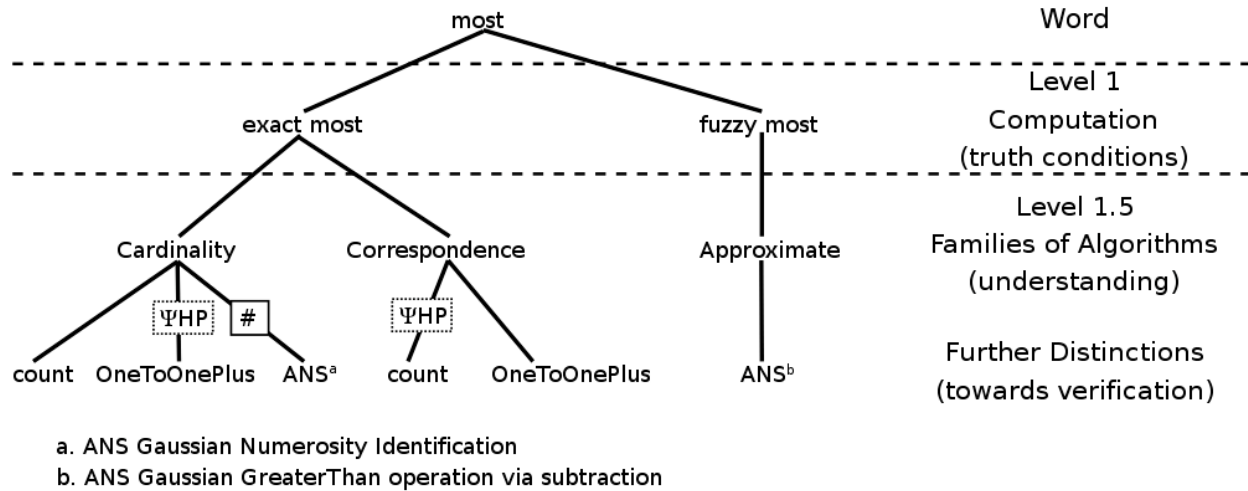
24

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 correspondence, 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. Step 1: present participants (at time scales that make counting impossible) with scenes in which it is easy to employ a OneToOnePlus 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 that 'most' is understood in terms of correspondence. Step 2: determine whether participants rely on the ANS to evaluate 'most'; and if yes, consider the possible mappings from meanings to this system.

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 OneToOnePlus strategy. A null result of this manipulation argues against a OneToOnePlus specification of what 'most' means. More strongly, we looked for positive evidence of participants relying on representations of the ANS. For as noted above, if such representations are invoked to answer a 'most'-query, this suggests that the query was not *understood as* one whose answer is the answer determined by a OneToOnePlus algorithm. 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.

6. 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, 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.
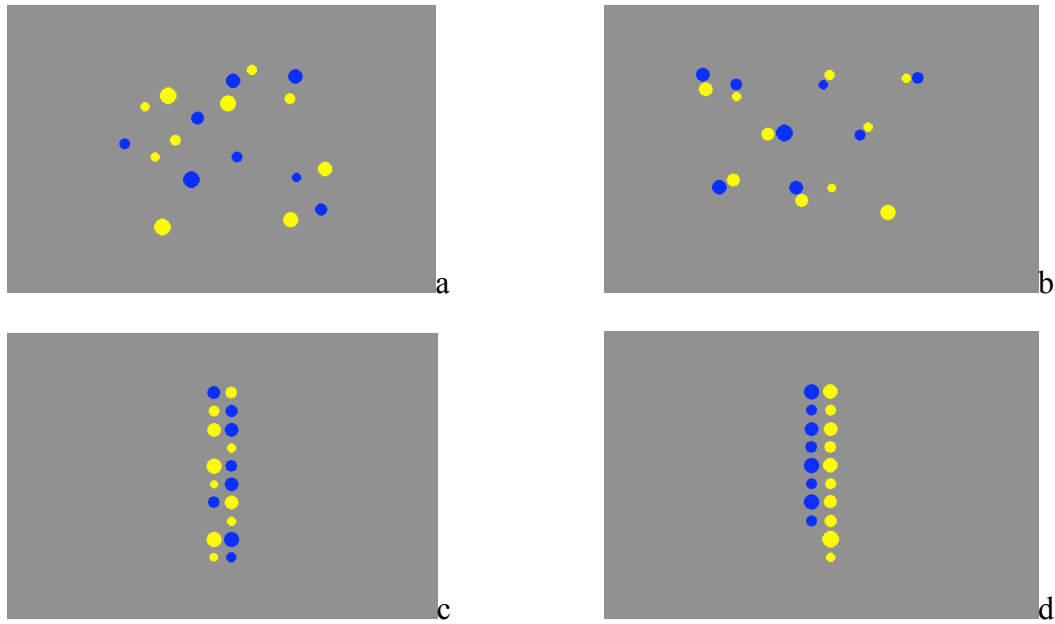
Figure 6

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

28

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 ($p = .277$), 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 OneToOnePlus algorithm, performance on Scattered Random trials patterned with performance on Scattered Pairs and Column Pairs Mixed, with percent correct declining as a function of 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

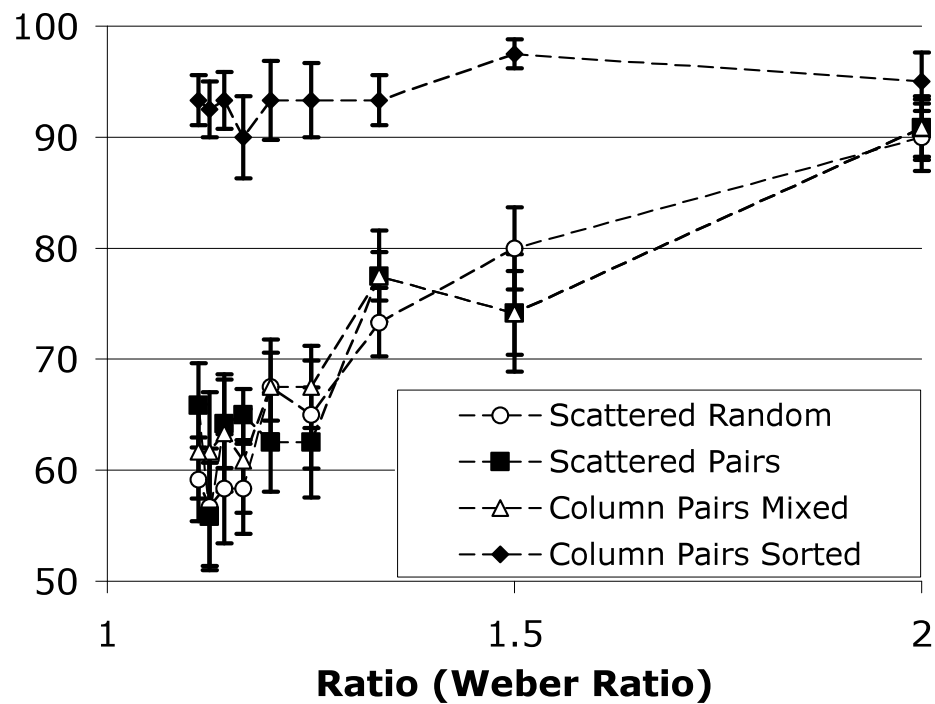| Trial Types	F	p
Scattered Random-Scattered Pairs	.216	.651
Scattered Random-Column Pairs Mixed	.446	.518
Scattered Pairs-Column Pairs Mixed	.127	.728
Column Pairs Sorted- Scattered Random	152.17	.0001
Column Pairs Sorted- Scattered Pairs	193.89	.0001
Column Pairs Sorted- Column Pairs Mixed	131.66	.0001

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 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:

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

The one free parameter in this equation is the 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 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 Weber fraction on these trial types suggests that participants relied on the representations of the Approximate Number System to evaluate 'most'.
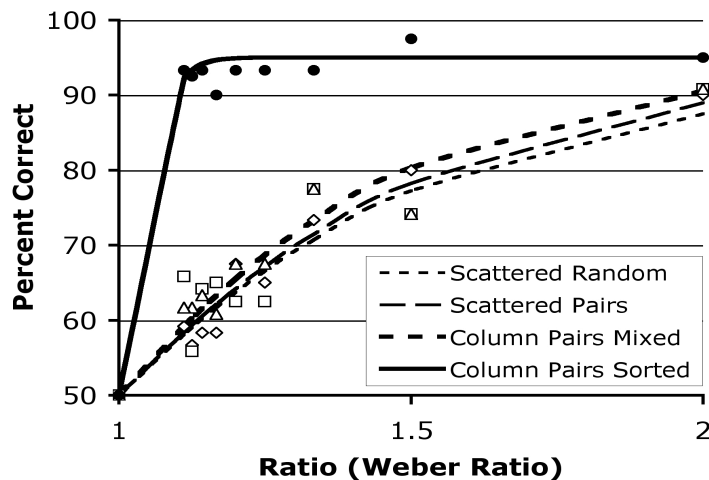

Figure 8

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 Weber fraction is expected to be approximately .14 for adults in number discrimination tasks and 'more'-tasks (Pica et al, 2005; Halberda & Feigenson, 2008), 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 Weber fraction of approximately .3 for three trial types suggests that participants may be translating the representations of the ANS into

32

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.[15] Further work is needed to determine if this is the case.

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 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

---

[15] Another possibility is that participants used their ANS to evaluate the 'most'-claim, but also relied on a subtraction, #{dots} − #{yellows}, to determine the number of nonyellow dots. This would increase the variability, and could be the source of our observed Weber fraction of .3, as we discuss in Lidz. et. al. (under review).

using a OneToOnePlus strategy for verification. Rather, they are using column length as a proxy for number (the perceptual "lines" created by the blue and yellow columns).[16]

We manipulated the ease of applying a OneToOnePlus 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 cannot be used to implement a OneToOnePlus strategy. As discussed in the next section, these results suggest that (at least for numerate adults), the meaning of 'most' is *not* specified in terms of one-to-one correspondence. In the next section, we also consider some alternative diagnoses of our results, and a related further question: 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?

7. Other Possibilities

One might think that our participants understood 'most' in terms of a OneToOnePlus algorithm, but that for some reason, their responses were not affected by the manipulations designed to make a OneToOnePlus strategy easier. In particular, one might worry that 200 ms of exposure was not long enough to detect the relevant yellow-blue correspondences and the color of any remainders, and then use this information—at least not in a way that would make Scattered Pairs *easier* than

---

[16] This notion of using column length as a proxy for number brings up issues of meaning versus verification again. While the correlation of number and column length in the present case empowers this type of verification procedure, we do not believe that the perceived line lengths are supplying numerical values to the 'most' algorithm. Rather, the correspondence of the 'most' algorithm and a separate 'longer-line' algorithm are recognized by the system and the subjects then simply performs the 'longest-line' algorithm and gives its answer as an answer for a most question.

Scattered Random with respect to answering the relevant OneToOnePlus question. We have several responses.

In ongoing work (Halberda, et.al., in prep), we have obtained initial evidence that 150 ms is long enough for participants to detect the relevant correspondences and answer a question depending on the (uniform) color of the remainders. In which case, a OneToOnePlus verification procedure is possible for these subjects, even though they do not use it. But more importantly, our experiment here tells against analyzing 'most' in terms of one-to-one correspondence.

For whatever 'most' means, participants were clearly able to use their ANS system to evaluate the target sentence. This is striking, but not inexplicable, if 'most' is understood in terms of cardinality comparison. Independent evidence suggests that the ANS interfaces, somehow, with representations of precise cardinalities; and given a "waystation" of the sort described above, our data is easily accommodated under a meaning for 'most' based in cardinalities. On the other hand, if 'Most dots are yellow' is understood as the claim that some of the yellow dots correspond one to one with all of the nonyellow dots, then participants must somehow evaluate this claim by using information provided by their ANS. But the only plausible "interfacing" mechanisms make the OneToOnePlus hypothesis unattractive.

Prima facie, this hypothesis requires a mind that can do the following in 200 ms: generate a semantic representation of 'Most dots are yellow'; generate (presumably in parallel) an ANS representation of the relevant scene; and then exploit some strategy for using the latter representation to answer the question posed by the former. This presumably requires *two* waystations: one that "converts" ANS information into comparisons of cardinalities; and another, effectively implementing one direction of (HP),

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

that converts cardinality comparisons intto correspondence comparisons. On this view, the net result, would be that participants answer the OneToOnePlus question posed by 'Most dots are yellow', by using the ANS to compare cardinalities.

By itself, the ANS is of no use in evaluating claims concerning one-to-one correspondence, since ANS representations are not representations of particular entities. Approximations of collection-numerosities are useful, but not for determining whether each nonyellow dot has a corresponding yellow dot—at least not without a cognitive bridge like (HP). Since speakers can and do connect 'most' to the ANS system, at least for purposes of natural verification, analyzing 'most' in terms of one-to-one correspondence evidently requires the following ancillary hypothesis: speakers also link 'most' to appropriate cardinality representations, perhaps via tacit grasp of a principle like (HP). But this undercuts the motivations, scouted in sections one and three, for the hypothesized OneToOnePlus specification/procedure for 'most'.

In response, one might say that participants do *try* to use a OneToOnePlus algorithm, but that their performance is error-prone at **200**ms. An errorless OneToOnePlus algorithm would not show this decrement in performance. But an error prone OneToOnePlus 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. As the yellow dots and blue dots become closer in number, such errors would become increasingly relevant to performance, leading to decreasing accuracy. This correctly predicts decreasing performance as a function of decreasing Ratio. But 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

OneToOnePlus algorithm predicts a difference in performance between these two trials. Because the two possible errors described above for an error prone OneToOnePlus 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 OneToOnePlus procedure predicts an effect of increasing accuracy with increasing number of dots in the display when ratio is held constant (Cordes et al, 2001). 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 OneToOnePlus 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 OneToOnePlus 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[17].

---

[17] 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' 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

A related thought is that participants might understand 'Most of the dots are yellow' as implying that the number of yellow dots is *significantly greater than* (and not merely greater than) the number of nonyellow dots. Perhaps 'most' requires 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 ($\approx$ 52% yellow) might *not* make 'Most of the dots are yellow' true, even if situations with 14 yellow dots and 9 blue dots ($\approx$ 60% yellow) would. This hypothesis predicts a decrease in participants' willingness to assent to 'Most of the dots are yellow' as 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 OneToOnePlus. If participants understood 'most' in this way, our observed result of a decrease in performance with decreasing Ratio could not be taken as evidence against OneToOnePlus. 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 our experiment , in which participants relied on the length of the

---

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.

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 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 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 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 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 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.

8. Conluding Remarks

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 (OneToOnePlus) 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 OneToOnePlus 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'.

At the end of section four, we briefly noted that this second finding is compatible with a scenario we set aside for present purposes. Instead of (5a), the possibility discussed here, perhaps the correct cardinality procedure/specification for 'most is better indicated with (5a');

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

(5a') Exceeds(G-ANS{x: Dot(x) & Yellow(x)}, G-ANS{x: Dot(x) & ¬Yellow(x)})

where 'G-ANS' signifies a mapping from sets to (not cardinalities, but rather) Gaussians represented by the ANS, and 'Exceeds' signifies a relation between Gaussians (each of which has a mean, mode, and standard deviation). Perhaps instead of using the ANS as a mere tool for judging whether one cardinality is greater than another, speakers specify the meaning of 'most' in terms of ANS-representations that exhibit an apparent order (see note 18).

Even if (5a) and (5a') are truth-conditionally equivalent—say because the relevant notion of order necessarily covaries with the means/medians/modes of the Gaussians—the specifications differ. On the latter view, a 'most'-claim concerns the order of the Gaussians corresponding to the yellow dots and nonyellow dots in question. And at least in principle, one can imagine minds that specify the meaning of 'most' in this way, even if they cannot (count or) represent cardinalities as such. Further experiments (Hunter, Lidz, Pietroski, and Halberda, in prep) are required to adjudicate between these alternatives.

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' (i.e., 'more' might be understood

as Cardinality-more or as OneToOnePlus-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 OneToOnePlus with those, or something else? To the best of our knowledge, these issues remain unsettled by the literature on 'more'. In any case, our participants did not report translating the target question this into a 'more'-question, and they seemed to engage with the task as presented. But in ongoing work (Lidz, Halberda, Pietroski & Hunter, in prep.), 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. Given dots of several colors, 'more'-judgments and 'most'-judgments can be teased apart.

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 truth-conditional contribution. Given such equivalences, choosing among alternatives requires appeal to other 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.

Davidson, D. (1974). Belief and the Basis of Meaning. *Synthese* 27: 309-23.

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.

Dummett, M. (1973): *Frege: Philosophy of Language*. London: Duckworth.

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. (2008). 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.*

Halberda, J., Lidz, J., Pietroski, P. and Hunter, T. (in prep), 'Language and number: Towards a

psychosemantics for natural language quantifiers'. Talk at 4th Hopkins Workshop on

Language. Oct 12-14, Baltimore, MD.

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

Horty, J. (2007). Frege on definitions: A case study of semantic content. Oxford: OUP.

Hunter, T., Lidz, J., Pietroski, P, and Halberda, J. (forthcoming). Beyond Truth Conditions: The Meaning of 'Most'. *SALT* proceedings.

Hunter, T., Lidz, J., Pietroski, P, and Halberda, J. (in prep).

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

Le Corre, M. & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition, 105,* 395-438.

Leslie, A., Gallistel, C. and Gelman, R. (2008). The generative basis of natural number concepts. *Trends in Cognitive Science* 12: 213-28.

Lidz, Hunter, Pietroski, and Halberda (under review).

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.

--(1992). *A Study of Concepts.* Cambridge, MA: MIT Press.

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.

Quine, W. (1960). *Word and Object*. Cambridge, MA: MIT Press.

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.