Interface Transparency and the Psychosemantics of *most*

Jeffrey Lidz, *University of Maryland*

Justin Halberda, *Johns Hopkins University*

Paul Pietroski, *University of Maryland*

Tim Hunter, *University of Maryland*

Contact Info:
Jeffrey Lidz
Department of Linguistics
University of Maryland
Marie Mount Hall
College Park, MD 20742 USA
Email: jlidz@umd.edu
Phone: +1 301-405-8220
Fax: +1 301-405-5xxx

Word Count Text:

Word Count Method and Figure Legends:

Title Character Count:

Keywords: analog magnitude, approximate number system, semantics-cognition interface, number, quantification, mathematics, most, language processing, language development.

Abstract

This paper proposes and defends an Interface Transparency Thesis concerning how linguistic meanings are related to the cognitive systems that are used to evaluate sentences for truth/falsity: a declarative sentence is semantically associated with a canonical procedure for determining its truth value (cf. Dummett 1973, Horty 2007); and while this procedure need not be used as a verification strategy, competent speakers are biased towards strategies that directly reflect canonical specifications of truth conditions. Evidence in favor of this hypothesis comes from a psycholinguistic experiment examining adult judgments concerning ‘Most of the dots are blue’.

This sentence is true if and only if the number of blue dots exceeds the number of nonblue dots. But this leaves many issues unsettled—e.g., how the second cardinality is specified for purposes of understanding and/or verification: via the *non*blue things, given a restriction to the dots, as in ‘$|\text{Dot}(x)\ \&\ {\sim}\text{Blue}(x)|$’; via the *blue* things, given the same restriction, and *subtraction* from the *number* of dots, as in ‘$|\text{Dot}(x)| - |\text{Dot}(x)\ \&\ \text{Blue}(x)|$’; etc. We obtained evidence in favor of the second hypothesis. Participants saw displays of between 2 and 5 colors of dots for 150ms, with between 5 and 17 dots per color. Ratios of blue to non-blue dots ranged between 1:2 and 7:8, with half the trials containing more blue dots and half containing more nonblues. Results indicated use of the Approximate Number System (Dehaene 1997) in verification, with accuracy *un*affected by the number of colors. It is independently known that for dots of up to 3 colors, the size of the total array is automatically computed. Given this fact and our result, we argue that the number of nonblue dots was specified via subtraction, and that sentences of the form ‘Most Δs are β’ have corresponding meanings, with truth conditions canonically specified along the following lines: $|\Delta(x)\ \&\ \beta(x)| > |\Delta(x)| - |\Delta(x)\ \&\ \beta(x)|$.

Where does meaning make contact with the rest of cognition?

Theories of meaning aim to specify the semantic properties of expressions. It is not obvious what these properties are. But traditionally, theories have been responsive to two basic concerns. First, a semantic theory for a natural language L is often said to be "empirically adequate" to the extent that the theory associates declarative *sentences* of L with *truth conditions* in accord with speakers' intuitions.[1] Second, such a theory must be *compositional*, at least in the following sense: the theory assigns "atomic" semantic properties to finitely many expressions of L; and for every other expression, its semantic properties are somehow determined by its constituents and their arrangement. In short, the idea is that a semantic theory should compositionally associate sentences with truth conditions.[2] Satisfying this requirement, even for a single language, remains a goal. But there has been progress, with many insights gained.

The catch, as every semanticist knows, is that given *one* compositional specification of truth conditions—say, for sentences of the form 'Most Δs are βs'—it is often easy to construct others. Even given various assumptions about the relevant syntax and its semantic role, there

[1] Cp. Davidson (1967) and Montague (1970), each of whom was inspired by Tarski's (1944) specification of a "materially adequate" notion of truth for certain invented languages. This leaves room for a pragmatics/semantics distinction, while allowing the use of model-theoretic techniques in describing entailments that competent speakers recognize.

[2] This is compatible with views according to which truth conditions are unstructured. (They might be conditional assignments of truth-values, or perhaps functions from possible worlds to truth-values.) But at least to first approximation: whatever "meanings" get assigned to sentences, they presumably *determine truth values* given the nonlinguistic facts, while being *determined by the constituent morphemes* given the relevant syntax.

may be many *truth-conditionally equivalent* compositional representations of the semantic properties exhibited by the expressions of a given language. Among theorists, there is broad agreement that not all such representations are equally good as proposals about how competent speakers *understand* expressions.[3] Put another way, many theorists suspect that sentential meanings are individuated more finely than truth conditions, and that distinct *formal specifications of* truth conditions can be suggestive of empirically distinguishable psychological hypotheses. But justifying specific proposals requires appeal to additional sources of evidence.

In this paper, we suggest and defend one such source: the interface between linguistic expressions and the extralinguistic cognitive systems that provide the information used, in contexts, to evaluate these expressions for truth/falsity. More narrowly, one can gain insight into the meaning of the determiner 'most' by examining how sentences like 'Most dots are blue' interface with the visual system. We argue that at least in this particular case, the sentential meaning constrains how the visual system can be used to evaluate the sentence, even to the point of blocking computations *native to the visual system* that would allow for more precise calculations.

Extending other work, our conclusion is that competent speakers associate sentences with *canonical specifications of* truth conditions, and that these specifications provide default verification procedures. From this perspective, examining how a sentence constrains its verification provides information about how speakers specify the truth condition in question.

---

[3] Evans (1981) suggested the potential relevance of many considerations, including aphasias; see also Davies (1987), Peacocke (1986), and Chomsky's (1986) E-language/I-language distinction, echoing Marr (1982) and Church (1941), who distinguished functions (in extension) from ways of computing them.

More generally, our data support an Interface Transparency Thesis, according to which speakers exhibit a bias towards verification procedures that employ the operations represented in the canonical specifications of truth conditions. This leads to substantive predictions, because a verification procedure of this sort may require (noisy) input representations that lead to *less* accuracy in judgment, compared with an alternative strategy that is cognitively available to speakers. To foreshadow: if speakers verify 'Most dots are blue' by computing the number of blue dots to *the result of subtracting* the number of blue dots from the number of dots—as suggested by the formal specification '|Dot(x) & Blue(x)| > |Dot(x)| - |Dot(x) & Blue(x)|'—this might lead to predictable *in*accuracies in judgment, thereby confirming the hypothesis that speakers employ the operation of cardinality subtraction as part of the default verification strategy.

There is nothing new in the idea that grammars (as internalized procedures) generate objects compatible with extralinguistic systems of perception, action and cognition. From the earliest days of generative phonology, linguists have been concerned with the relation between phonological, articulatory and acoustic properties of speech (Jakobson, Fant & Halle 1952; Liberman et al 1967, Stevens 1972), asking about the degree to which phonological properties are constrained by the extralinguistic systems of articulation and audition, both in the acquired grammar (Liberman & Mattingly 1985, Halle 1999, Poeppel, et al. 2008) and in the acquisition process (Kuhl 1993, Werker 1995, Jusczyk 1997). But until more recently, the tradition in natural language semantics has been to focus on relations that expressions bear to entities in an idealized model of the world that speakers talk about, as opposed to intensive investigation of the constraints placed on the grammar and its acquisition by extralinguistic systems of thought.

As a long notable exception, Jackendoff (1983, 1990, 2002) has usefully illustrated how theorists can draw conclusions about conceptual structures from linguistic data. Inferences in the opposite direction have been harder to come by (but see Landau and Jackendoff 1993 for one attempt). This difficulty derives in part from the fact that one cannot be sure which conceptual systems interface with the language faculty, and in part from the fact that the especially relevant cognitive subsystems have not been adequately described.

In the current paper, we focus on sentences containing the quantificational determiner *most*, as a case study of the relation between the cognitive and linguistic representations of quantification, comparison and measurement. Quantificational expressions represent a valuable preparation for cognitive science as they have benefited from extensive study within multiple disciplines including linguistics, philosophy and psychology, making precise hypotheses about the interface between semantics and cognition possible.

Proportional quantifiers like *most* have long been of interest to linguists because they cannot be defined within first-order predicate logic and therefore provide a lower bound on the expressive capacity of natural language (Barwise and Cooper 1981). To accommodate this expressive capacity, Barwise and Cooper (1981) adopted Generalized Quantifier Theory (Mostkowski 1957), treating quantifiers as expressing relations between sets (see also Higginbotham and May 1981). For example, *most* can be treated as expressing a comparative relation between the cardinalities of two sets—or equivalently, as a function that maps each ordered pair of sets (X, Y) to a truth value as in (1a).[4] Correlatively, sentence (1b) is true iff the broken crayons outnumber the crayons that are not broken.

---

[4] Note that while '>' signifies a relation between cardinalities, '–' does not. Here, '–' signifies set-subtraction, not cardinality-subtraction: '$|A - B|$' is equivalent to '$|\{x: (x \in A) \, \& \sim(x \in B)|$'.

(1) a. $\text{MOST}(X, Y) = \text{TRUE iff } |A \cap B| > |A - B|, \text{ otherwise FALSE}$
b. Most of the crayons are broken.

While Generalized Quantifier Theory (GQT) provides a framework for unifying proportional determiners like *most* with other quantifiers, this theory is silent with respect to a number of alternative expressions of the meaning of a determiner that are alike in both truth-conditional import and compositional properties (Hackl 2008, Pietroski et al 2008). While GQT requires that proportional determiners express relations between sets, each such relation can be described in many truth-conditionally equivalent ways; and each such description is equally good, so far as GQT is concerned. But presumably, speakers who represent the psychological content that makes understanding possible in some specific format and the particular algorithm that implements a linguistic meaning must be able to interface with this format. In what follows we show how understanding the psychology of non-linguistic quantification constrains and informs the proper treatment of proportional quantification in natural language.

We consider here only one quantifier, *most*, which appears to get its content, at least in part, from nonlinguistic number representations (Pietroski et al 2008). Through psychophysical experimentation, we show that the particular way in which the cognitive representations of quantity are deployed in the process of understanding a sentence containing *most* provides evidence about the lexical semantics of the quantifier itself.

Truth-Conditionally Equivalent Alternatives

Treating a proportional quantifier like *most* as expressing a relation between two sets leaves room for a multitude of possible formulations of this relation. For example, in Pietroski et al (2008), we observed that the meaning of a sentence like (2) might equivalently be represented in two ways, given in (3).

(2) Most of the dots are blue.

(3) a. >(|DOT & BLUE|, |DOT – BLUE|)

b. OneToOnePlus({DOT & BLUE}, {DOT – BLUE})

The relation in (3a) expresses the familiar comparative relation over cardinalities. The symbol ">" expresses a relation that holds between two cardinalities, with cardinality being a property of sets. Thus, (3a) means that the cardinality of the set of blue dots exceeds the cardinality of the set of non-blue dots. The relation OneToOnePlus, found in (3b), holds between two sets A and B just in case there is a one-to-one relation that holds between B and a proper subset of A. Thus, (3b) expresses the idea that if you paired each nonblue dot with exactly one blue dot, there would be at least one blue dot remaining.

These representations are alike in treating *most* as expressing a relation between the set of blue dots and the set of nonblue dots. Thus, from the perspective of compositionality, these alternatives are equivalent. In addition, they are equivalent at the level of truth conditional import, since both will be *true* when the blue dots outnumber the nonblue dots and *false* otherwise.

Despite this equivalence, however, the two representations imply different algorithms for deriving the same truth condition. Both express a relation between two sets (the blue dots and the nonblue dots), but only one of them invokes the notion of cardinality in the relation. Both require that there be more blue dots than nonblue dots, but only one of them makes this comparision an explicit part of the algorithm for determining the truth condition. In short, both treat the word *most* as expressing a relation between two sets, both are identical in the function from contexts to truth-values, but they differ in the conceptual resources out of which this function is built.

Hackl (2008) highlights the difference between the truth-conditional import of an expression and the algorithm for building up the semantic representation, also in the domain of *most*. He points out that there are alternative formulations of which sets are compared even if *most* is treated as a relation over cardinalities. Thus, in addition to (3a-b), one might represent the meaning of (2) as in (4).

(4) >(|DOT & BLUE|, ½ |DOT|)

As above, this expression entails that the blue dots outnumber the nonblue dots. If the blue dots make up more than half of the total number of dots, then it must be the case that the blues outnumber the nonblues.

Given the availability of alternatives (and there are obviously many more truth-conditionally equivalent variants yet to be discussed), one wants to know whether there is a fact of the matter. Are these alternatives merely notational variants about which it makes no sense to ask which is correct (much as it makes no sense whether temperature is properly represented in Farenheit or Celsius)? Or, can these alternatives be taken as alternative psychological hypotheses about speakers of English?

We pursue the latter option, taking the position that the meaning of a sentence is something more than a compositionally determined function from worlds to truth-conditions. We argue that alternative representations of a truth-condition are legitimate psychological alternatives about the computations involved in determining that truth-condition. But, as noted above, given the truth-conditional equivalencies, the algorithms for building the meaning of a sentence must be investigated by means other than the means of standard linguistic semantics.

On Verification Procedures

It is important at this point to emphasize that the differences we have been talking about reside in algorithms for building up truth-conditions compositionally, and not algorithms for determining truth in a context. For any truth condition, however computed, there will (at least in principle) be many methods for determining whether the condition obtains. If you want to know whether (5a) is true, you might check some rabbits, or check a website.

(5) a. Rabbits are furry
b. Chicago has great architecture
c. Most of the dots are blue
d. La niege est blanche

If you want to know whether (5b) is true, you might go to Chicago and look around, or you might read a book. If you want to know whether (5c) is true, you might count (if you have the time and opportunity), you might estimate the relevant cardinalities, or you might just ask someone next to you, especially if you are color blind. And of course, if you want to know whether (5d) is true, there are ways of finding out without even understanding the sentence (like asking someone who speaks French). But when a speaker understands a (novel) sentence and judges it to be true or false in a given context, the speaker presumably does at least two things: compositionally determine a truth condition; and, determine whether that condition obtains in the context. And at least typically, the latter presupposes the former.

You can reliably assess the truth of a sentence by asking your neighbor only if you can treat your neighbors' response as a reliable indicator of whether the truth condition associated with that sentence actually obtains. And in the general case, this requires that you know which truth condition this is. But suppose, for illustration, that you understand (2) as in (3a). Then in relying on your neighbor, you are effectively relying on your neighbor to tell you whether the number of blue dots is greater than the number of nonblue dots. But this *verification procedure*, for figuring out whether (2) is true or false, does *not* proceed as follows: figure out the number of blue dots, figure out the number of nonblue dots, and figure out if the first number is bigger. Your neighbor may or may not go through this procedure; but you don't. So understanding (2) as in (3a) *does not* commit you to going through this procedure, or through any particular procedure, to determine whether (2) is true. In this sense, verification procedures are independent of the algorithms that produce truth conditions. But of course, the condition one seeks to "verify" is typically determined by one's procedure for determining truth conditions, i.e., by the compositional structure of the sentence.

However, procedures for determining truth conditions *do* provide verification procedures in one perfectly fine sense. If you understand (2) as in (3a), then you presumably know that one *could* determine the truth or falsity of (2) by determining and comparing the relevant cardinalities. In a given context, you might not be able to determine the truth or falsity of (2) in this way; and perhaps in practice, no real person could. Maybe the relevant dots are *very* far away, locked in a big black box. More generally, there is no guarantee that the verification procedure suggested by a procedure for determining a truth condition will be practical. But sometimes it will be.When conditions are favorable, one can figure out if (2) is true by determining and comparing two cardinalities.

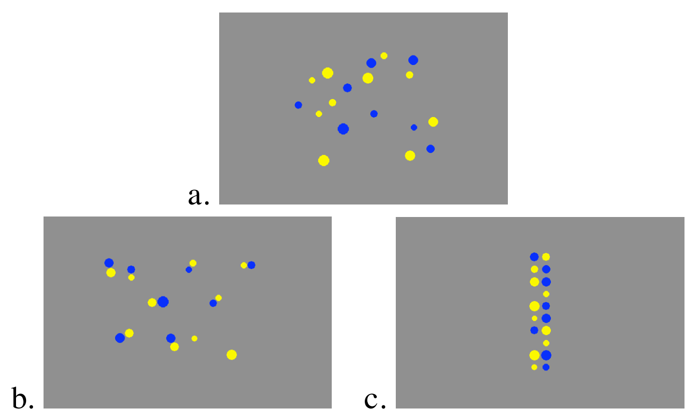
Indeed, in the experimentation presented below, we present evidence for the following hypothesis:

> Interface Transparency Thesis (ITT): the verification procedures employed in sentence understanding are biased towards those algorithms that directly compute the relations expressed in the meaning.

The ITT holds that a sentence meaning is not verification independent, but rather provides an instruction to the interface systems concerning what information to gather in order to verify the sentence. At least for the logical vocabulary, the meaning bearing units that contribute to a truth condition may be related to verification procedures in the following way: a lexical item is understood in terms of a certain mental representation; and this representation suggests a verification procedure that speakers can execute in contexts where conditions are favorable. Let us call this suggested verification procedure the *canonical specification* of the meaning. Below we test the ITT by pitting the putative bias for transparency against the native computations of the visual system. What we will see is that a sentence meaning can cause the visual system to overlook information that it otherwise computes automatically, thus supporting the ITT. Before turning to our experimentation, however, it is worth reviewing some earlier results that take a similar view on the relation between verification and meaning.

Most: Prior Results

In Pietroski et al (2008), we put people in a range of situations that differed in their amenability to a verification procedure for *most* stated in terms of OneToOnePlus, as in (3b). Consider the displays in (6). Each contains 10 yellow and 8 blue dots, but they differ in the degree to which they invite pairing the dots. In (6a), the dots are scattered randomly on the screen. In (6b), they are scattered, but in pairs such that the only singleton dots come from the larger of the two sets. In (6c), the dots are arranged in two columns and 10 rows, with each row consisting either of a pair of yellow and blue dots, or only a singleton yellow dot.



(6)

We flashed a series of such displays on a computer screen for 200ms and asked people to determine on each trial whether the sentence (2) was true. We found that despite the fact that the displays varied in their suitability to a verification procedure stated in terms of one-to-one correspondence, like (3b), accuracy was unaffected by this manipulation. We thus concluded that the compositional representation associated with the sentence does not include a one-to-one correspondence relation between the blue and non-blue dots. If it did, then scenarios inviting the use of a verification procedure in exactly those terms would have invoked such a procedure and we would have found differences in response patterns across conditions.

More positively, our data also provided evidence that subjects were making use of an approximate representation of numerosity in order to answer the *most* question. Support for this idea came from the fact that adult responses to the *most* question showed the behavioral signature of the Approximate Number System (ANS) (Dehaene 1997, Feigenson et al., 2004). The ANS generates an approximate representation of the number of items in a set. This representation adheres to Weber's law (the discriminability of two quantities is a function of their ratio). The ANS is an evolutionarily ancient piece of cognitive machinery that is shared throughout the animal kingdom and does not require explicit training with number in order to develop (*for review see* Feigenson et al, 2004). The adult subjects in Pietroski et al. (2008) were shown a display of yellow and blue dots for 200ms and asked to determine if most of the dots were yellow. We found that the probability of answering yes or no was a function of the ratio of the numerosity of the set of yellow dots to the numerosity of the set of non-yellow (blue) dots. In other words, participants were better able to determine whether most of the dots were yellow as the ratio of yellow dots to blue dots increased. This increase was well fit by a psychophysical model that relies solely on the representations of the Approximate Number System (Pica et al. 2006; Halberda, Mazzocco & Feigenson 2008).

That the responses for *most* judgements were fit by this model supports the view that representations of the Approximate Number System were involved in the verification procedure, and hence that, at least under some conditions, the numerical content required to verify the meaning of a sentence containing *most* is provided by the Approximate Number System. It is important to realize, however, that implicating the approxmiate number system in the verification procedure is not the same as implicating this system in the semantic procedure that generates truth conditions, i.e., the mental representation of the syntax and semantics of the expression. More generally, evidence that people do use a particular verification procedure is not evidence that the meaning forces them to. In this case, the fact that people used the ANS to gather information about cardinality does not provide evidence that the relation between cardinalities that is expressed by *most* is represented in analog terms, only that in 200ms the ANS can provide information that this relation demands.

Putting aside the question of whether the cardinality comparisons for *most* involve precise or approximate cardinalities, recall the distinction highlighted in Hackl (2008):

(7) a. $\text{most}(A)(B) = 1 \text{ iff } |A \cap B| > \frac{1}{2}|A|$

b. $\text{most}(A)(B) = 1 \text{ iff } |A \cap B| > |A - B|$

Both of these expressions are compatible with the finding that determining the meaning of *most* involves a comparison of cardinalities. In order to distinguish these alternatives, Hackl (2008) asked people to determine the meaning of sentences like (8a) and (8b) in an experimental paradigm he called “self-paced counting”.

(8) a. Most of the dots are blue.

b. More than half of the dots are blue.

In the self-paced counting paradigm, participants see a series of empty circles on the screen. The participants then press a button, causing some of the dots become colored in (either red or blue). At the next button press, those return to being empty circles and a subsequent subset becomes colored in. This continues until the colors of all of the dots have been revealed and the participant says whether the test sentence was true or false.

Hackl found that reaction times to sentences like (8a) were significantly faster than reaction times to sentences like (8b). He reasoned therefore that the meaning of (8a) must not be equivalent to the meaning of (8b). Since (7a) is transparently the meaning of (8a), it follows that it is not the meaning of (8b).

Prior literature providing psychological evidence concerning the lexical representation of *most* thus leads to several conclusions. First, the meaning of *most* expresses a relation between cardinalities, but not a relation involving one-to-one pairing of individuals in a set. Second, this relation does not involve a comparison of one subset with ½ of the total. Third, at least for the purposes of verification, the cardinalities to be compared are provided by the approximate number system.

While these are important steps forward, additional questions remain. Rejecting (3b) and (7a) as the meaning of *most* does not entail that we should accept (3a)/(7b). There are more alternatives on the table. We thus seek to determine positively which cardinalities get compared in determining the truth of a sentence containing *most*.

On the face of things, the question of which cardinalities are compared in these experiments seems obtuse. In both Pietroski et al (2008) and Hackl (2008), participants were repeatedly presented with dots of 2 colors (e.g., yellow & blue) and asked each time whether most of the dots were blue. Quite obviously, they were comparing the number of blue dots to the number of yellow dots. Such a conclusion might be too rash, however. If people were, in fact, comparing the number of blue dots to the number of yellow dots, then they were computing the value of sets that are not explicitly represented in the sentence they were asked to judge. Let us consider two ways of representing the meaning of the target sentence (9a), given in (9b) and (9c):

(9) a. Most of the dots are blue

b. |DOT & BLUE| > |DOT & -BLUE|

c. |DOT & BLUE| > |DOT| – |DOT & BLUE|

The representation in (9a) treats the 2nd argument of the > relation as the intersection of two sets, one positively defined (the dots) and the other negatively defined (the non-blue things). The representation in (9c) is different. It treats the 2nd argument of the > relation as being computed by subtraction. That is, the cardinality of non-blue dots is determined by subtracting the blue dots from the set of dots.

There are two important things to note about these representations. First, neither of them contains any mention of the property "yellow". Thus, if people were answering the question, "*are most of the dots blue?"* by comparing the number of blue dots to the number of yellow dots, then they were doing something of a transformation on the meaning of the expression. They would have to have gone through an implicit chain of reasoning something like the following. "I am being asked to compare the size of the set of blue dots against the size of the set of nonblue dots. In this context the nonblue dots are always yellow. Thus, I should compare the size of the set of blue dots to the size of the set of yellow dots." In other words, given a meaning like either (9b) or (9c), people would have transformed the meaning into a verification procedure along the lines of (10):

(10) |DOT&BLUE| > |DOT&YELLOW|

In essence, this would involve resolving the reference of the 2nd argument of > prior to making the numerical comparison. Given that a comparison of the number of blue dots to the number of yellow dots requires a transformation of the 2nd argument, it becomes plausible to imagine that people were not, in fact, comparing the numerosities of the blue and yellow dots.

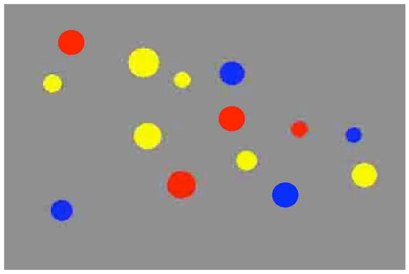
Indeed, if we were to find that the computation involved an explicit comparison of the numerosities of the blue and yellow dots, that would provide evidence against the Interface Transparency Thesis, since this thesis demands that the verification procedure compute the canonical specification of the meaning without such a transformation.

Now, the verification procedure does not provide direct evidence about the mental representation of the sentence's meaning (i.e., the semantic constituents out of which the truth condition is constructed). On the other hand, verification does provide a window into meaning, especially if the ITT is true. To see how, suppose that we knew that selecting and enumerating the yellow dots was a rapid and automatic computation of the visual system. And suppose further that we could find evidence that the verification procedure did not make use of this information. Such evidence might then be taken as evidence in favor of the ITT. That is, if the relevant interface system fails to use information that it automatically computes and that would in principle be useful in verification, then something, most plausibly the semantic representation, must be responsible for such a failure. It is precisely this argument that we now undertake to pursue.

Interface constraints imposed by the visual system

Imagine that a listener was shown a briefly flashed display of dots of many colors (Figure 1) and was asked to assess whether (11) is true of the display.

Figure 1



(11) Most of the dots are blue

Now recall the two alternatives presented above for the representation of the second argument of the > relation:

(12) a. $|\text{DOT \& BLUE}| > |\text{DOT \& -BLUE}|$

b. $|\text{DOT \& BLUE}| > |\text{DOT}| - |\text{DOT \& BLUE}|$

The meaning expressed in (12a) invites a verification procedure that attends to and enumerates the blue dots and the non-blue dots, comparing the numerosity of the former with that of the latter. The meaning expressed in (12b) invites a verification procedure that attends to and enumerates the set of dots and the set of blue dots, subtracting the latter from the former to construct the second argument of the > relation, and then comparing the result to the numerosity of the set of blue dots. In short, the difference between the two verification procedures lies in whether the set of nonblue dots is **selected,** with a subsequent step of estimating its cardinality, as in (12a), or whether this cardinality is **computed**, as in (12b).

We can therefore ask whether it is psychologically possible to directly select and enumerate both the blue dots (as both computations would require) and the nonblue dots (as required only for 12a). Even without briefly flashing the array, the reader can likely experience that selecting only the blue dots from among all of the dots is easy. Research on adults' ability to search for a colored item among colored distractors has shown this to be the case; *blue*, and all other categorizable colors, works as an early visual feature that can be found very quickly in a visual scene when the distractors are of saliently different colors, as they are in Figure 1 (Wolfe 1998; Halberda et al, 2006). But similar research also reveals that a set defined by a disjunctive combination of early visual features (e.g. dots that are either yellow OR red) is *not* easily selectable. That is, adults are unable to rapidly search all items in an array in order to find all the items that are either yellow or red (Wolfe 1998, Treisman & Gormican 1988, Treisman & Souther 1985). This calls into question the viability of having the meaning in (12a) map directly onto a verification procedure which requires listeners to directly attend and enumerate both the blue and the non-blue dots for purposes of ordinal comparison. Because the non-blue dots are a heterogenous set, they can not be attended directly. Moreover, building up the non-blue dots by constructing a disjunctive combination of each non-blue set is also not a straightforward visual computation. Listeners simply would not be able to directly attend the heterogeneous set of non-blue dots.

But, looking at Figure 1, it seems that we *can* assess whether most of the dots are blue, and so the question becomes (i) how we are accomplishing this and (ii) whether (12b) provides a more natural verification procedure. Additional evidence from the psychological literature is helpful in this regard.

Halberda and colleagues (2006) have demonstrated that adults can use the Approximate Number System to estimate the cardinality of up to three sets in parallel. On each trial in Halberda et al (2006), participants were shown a brief flash that contained from 1 to 6 colors of dots randomly scattered on a black background, similar to Figure 1. Either before or after the flash, participants were asked to approximately enumerate only one of the sets (either the superset of all dots irrespective of color, or a particular color subset). On a "Probe After" trial, where subjects did not know which set to report until after the flash had gone, the most likely strategy is to enumerate as many sets as possible and hope that one of those sets would be the one asked. By comparing performance on Probe After to Probe Before trials, Halberda et al (2006) were able to estimate how many sets adults could enumerate from a single flash. Results suggested that adults *always* attend and enumerate the superset of all dots. In addition to the superset, adults could also attend and enumerate some of the color subsets on multi-color trials. The typical adult appeared to enumerate the superset of all dots and two of the color subsets, but no more. For example, shown the flash depicted in Figure 1, a typical adult would know that there had been approximately 14 total dots, and perhaps that there had been approximately 4 red dots and approximately 6 yellow dots but nothing more.

That adults can enumerate multiple sets from a single flash using the Approximate Number System highlights the potential relevance of this system for verification procedures associated with natural language quantifiers like *most*. A meaning like (12a), translated directly into a verification procedure, is implausible because it involves selecting a heterogeneous set. However, this meaning invites the transformation in (13), wherein the set of non-blue dots is constructed by summing the cardinalities of each color subset comprising the non-blue dots.

(13) |DOT&BLUE| > |DOT&RED| + |DOT&YELLOW|

However, such a transformation would be useful only when the display contains no more than 3 colors, given Halberda et al.'s observation of a 3-set limit on early visual attention. That is, to verify this meaning would require the visual system to attend the color subset of blue dots, the color subset of red dots and the color subset of yellow dots. If there are only these three colors present in the array, then an addition of yellow and red dots would provide the listener with the number of non-blue dots, which could then be compared to the number of blue dots to yield a truth value. But, because adult humans appear to be limited to enumerating only up to three sets at once, this verification procedure, and hence the meaning in (12a), becomes less plausible as the number of color subsets increases.

A meaning like (12b), however, is straightforwardly verified with these resources, since the sets required for its verification (one color plus the superset) are easily and automatically attended by the visual system. Moreover, this meaning does not become less plausible as the number of color subsets increases.[5] That is, to verify a meaning like (12b) would require first enumerating the superset of all dots and the color subset of blue dots. The next step would involve a subtraction of these two values to calculate the number of non-blue dots. The final step would compare the number of blue dots to the number of non-blue dots to yield a truth value. Because only the superset and one color subset need be attended, the meaning in (12b), along with its associated verification procedure, is psychologically plausible, no matter how many color subsets there are, so long as it is possible to perform the subtraction and comparision computations.

In order to determine whether the canonical specification of the meaning associated with *most* is like (12a) or like (12b), we asked adult participants to verify whether most of the dots in an array were blue across many trials where we randomly varied the number of colors in the array. If participants verify *most* via the meaning expressed in (12a), then we expect accuracy to

---

[5] Halberda and colleagues (2006) found no reduction in enumeration accuracy for adults' ability to enumerate a color subset when the number of colors in the distractor subsets increased. Performance was the same for enumerating the yellow dots if there were no other colors present, yellow and blue dots, yellow blue and red dots or even yellow blue red green purple and cyan dots. Also, Halberda et al (2006) found no cost for estimating the cardinality of the superset as the number of colors in the stimulus increased. So, enumeration of the yellow dots and the superset appear to be unaffected by increasing the number of color subsets making the meaning expressed in 3b plausible as the number of sets increases and the meaning in 3c implausible.

decline as the number of colors in the array increases. On the other hand, if participants verify *most* via the meaning expressed in (12b), then we expect the number of colors to have no impact on their responses.

Experiment

We used a common visual identification paradigm to evaluate the underlying meaning for ‘most’.

Method

Participants

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

Materials and Apparatus

Each participant viewed 400 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 0.8 degrees of visual angle from a viewing distance of 50 cm.

Design and Procedure

On each trial, subjects saw a 150ms display containing dots of at least two colors and at most five colors (blue, yellow, red, cyan, magenta). Blue dots were present on every trial. Subjects were asked to answer the question "Are most of the dots blue?" for each trial. The number of dots of each color varied between 5 and 17. Whether the blue set represented more than half of the total number of dots (that is, whether the correct answer to "Are most of the dots blue?" was yes or no) was randomized. Subjects answered "yes" or "no" by pressing buttons on a keyboard. Within each trial type (i.e., 2-5 colors), the ratio of blue to non-blue dots varied between 5 possible ratios (1:2, 2:3, 3:4, 5:6, and 7:8). Within each of these ratio bins the blue set was the larger set on half of the trials.

Half of the trials for each trial type (2-5 colors) for each ratio bin were "dot size-controlled" trials on which, while individual dot sizes varied, but the size of the average blue dot was equal to the size of the average yellow dot, the average green dot and so on. On dot-size controlled trials the set with the larger number of dots would also have a larger total area on the screen (i.e. more total blue pixels when blue was the larger set). The other half of the trials were "area-controlled" trials in which individual dot sizes varied and the total amount of blue and non-blue pixels on screen was equated (i.e. smaller blue dots on average when blue was the larger set). On both dot 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.

All trials were randomly shuffled such that number of colors (2-5), correct answer (yes/no), ratio bin (1:2-7:8), and stimulus type (dot-size controlled, area controlled) varied randomly during the experiment.

Predictions

If subjects rely on the imprecise cardinality representations of the Approximate Number System (ANS) then accuracy should decline as a function of ratio, and should be well-fit by a psychophysical model of the ANS. With respect to the question of whether (12a) or (12b) underlies the meaning of *most*, we consider two hypotheses. First, if subjects determine the set of "non-blue" dots by determining the cardinality of each subset and then summing the non-blues together (algorithm 12a/13), we predict that subjects should succeed at the task when there are two and perhaps three colors on the screen but that performance should rapidly decline for higher numbers of colors (we will call this the "sum the non-blue" hypothesis). This prediction derives from the observation from Halberda et al (2006) that at rapid presentation rates, the visual system can accurately track a maximum of 3 sets. The second hypothesis, which we will call the "subtraction" hypothesis, holds that the computation of the meaning of *most* is identical across all trial types. The cardinality of the set of "non-blue" dots is determined by subtracting the cardinality of the focused set (the blue set) from the cardinality of the superset (the dots). Thus, since only two sets ever need to be selected by the visual system, the number of colors should have no impact on responses.

Results:

Results were entirely consistent with the subtraction hypothesis, suggesting that algorithm (12b) reflects the canonical specification of the meaning of *most* and that this algorithm relies on the representations of the Approximate Number System. There were no differences across trial types as a function of the number of colors in the display (Figure 2) and perfomance on every trial type was perfectly fit by a psychophysics model of the ANS (Table 1 and Figure 3).

A detailed description of the statistics and figures follows. Percent correct for each participant for each ratio was entered into a 4 Trial Type (2-, 3-, 4-, 5-Colors) X 2 Stimulus Type (dot size-controlled, area-controlled) X 5 Ratio Repeated Measures ANOVA. There was a significant effect of Ratio as subjects did better with easier ratios: $F(4, 44) = 109.092$, $p < .001$, a significant effect of Stimulus Type as subjects did slightly better on dot size-controlled than area-controlled trials: $F(1, 11) = 7.326$, $p < .05$; and most importantly there was no effect of Trial Type as subjects did equally well independent of the number of colors in the stimulus: $F(3, 33) = 0.276$, $p = .842$. Because the small but significant Stimulus Type effect does not bear on the inferences we make about the algorithms involved, we combined performance for each subject for each Ratio and each Trial Type for further analyses.

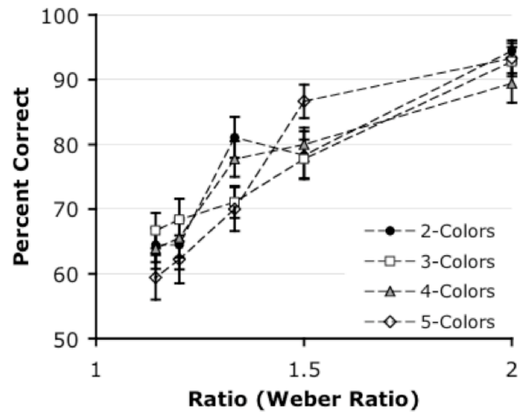


Figure 2

As can be seen in Figure 2, while performance declines as a function of Ratio, performance is the same independent of the number of colors in the array[6]. This supports the predictions of the subtraction hypothesis that on every trial type, irrespective of the number of colors in the display, subjects attend the superset of all dots and the focused set (blue dots), enumerate each and then perform a subtraction in order to calculate the number of non-blue dots before comparing the number of blue dots to the number of non-blue dots. Obviously, we are not suggesting that this subtraction is a conscious subtraction and we doubt that subjects are even aware of how they are figuring out what answer to give. The subtraction hypothesis, i.e., the meaning expressed in (12b), is meant to characterize the unconscious computations that underlie the meaning of *most* and allow it to interface with the rest of psychology.

The Approximate Number System is known to contain both the representational and computational machinery necessary to represent imprecise cardinalities, perform subtractions of these cardinalities, and make ordinal comparisons of these cardinalities (Whalen, Gallistel &

[6] Throughout the analyses, ratios will be displayed as the Weber Ratio between the two sets (Weber Ratio = bigger #/smaller #). This is important as it allows performance to be fit by a psychophysical model of the Approximate Number System.

Gelman 1999, Dehaene 1997, Feigenson, Dehaene & Spelke 2004, Brannon, Lutz & Cordes 2006). Thus, the ANS itself may be capable of implementing the entire algorithm expressed in (12b). A first step in evaluating whether this is the case is to see if performance on each Trial Type can be fit by a computational model of the ANS.

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 subjects is modeled for each Trial Type as a function of increasing Weber Ratio (larger set/smaller set, or n2/n1). Each numerosity is represented as a Gaussian random variable (i.e., X2 & X1) with means n2 & 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 that has a mean of n2-n1 and a standard deviation of $w\sqrt{n1^2+n2^2}$ (simply the difference of two Gaussian random variables). Percent correct is then equal to the area under the resulting Gaussian curve which is to the right of zero, computed as:

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

The one free parameter in this equation is the critical Weber Fraction (w). This parameter determines percent correct for every Weber Ratio (N1/N2). 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 non-blue dots would in fact be identical) were fit using this psychophysical model. As can be seen in Figure 3, the fits for all four Trial Types (2-5 Colors) fell directly on top of one another. Table 1 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 subjects' performance in the experimental task (R^2 values > .9). The critical Weber fraction on these trial types confirms our earlier result that participants rely on the representations of the Approximate Number System to evaluate *most*.

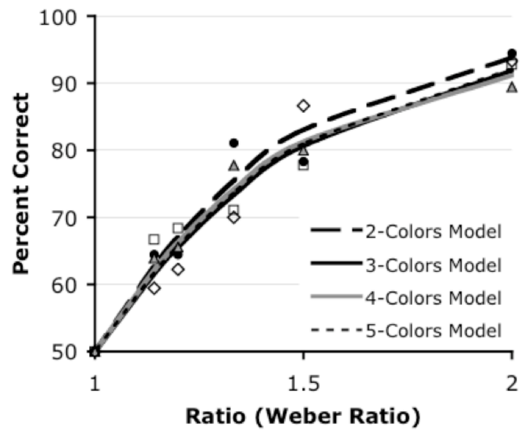


Figure 3

Table 1. Parameter estimates from psychophysical model

Trial Type	R^2	Critical Weber Fraction	Nearest Whole-Number Ratio
2-Colors	.9480	.290	3:4
3-Colors	.9586	.320	3:4
4-Colors	.9813	.283	3:4
5-Colors	.9625	.316	3:4

The Weber fraction is expected to be approximately .14 for adults in number discrimination tasks, i.e. *more*-tasks, (Pica et al, 2005) and range from .14- to .35 in adults when subjects are translating these representations into whole-number values (Halberda et al, 2006; Whalen et al, 1999). Our estimate of a Weber fraction of approximately .3 for all four trial types suggests that subjects may be translating the representations of the ANS into whole number values before evaluating *most* (*see also* Pietroski et al, 2008). That is, shown an array of 28 total dots, 16 of which are blue, these subjects may activate the ANS representations for 28 and 16, perform an ANS subtraction to represent the 12 non-blue dots, and translate the values 12 and 16 into whole-number estimates *twelve* and *sixteen* for purposes of evaluating *most*. Another possibility is that the entire computation is done within the ANS without ever translating into whole-number values. In such a model, the dual operations of subtraction and ordinal comparison may each contribute to determining the Weber Fraction. Further work will be necessary to tease these two possibilities apart.

General Discussion

We found no change in participants' ability to evaluate *most* as a function of the heterogeneity of a display. Rather, participants' performance at evaluating *most* for a wide range of ratios across all trial types was best fit by a model of the Approximate Number System whereby participants rely on a subtraction to compute the cardinalities of the sets to be compared. These results inform our understanding of how the meaning of *most* interfaces with the psychological mechanisms that provide numerical content, and lays the groundwork for further investigation of the interface between language and number.

More generally, our research addresses the relation between the units of meaning out of which truth conditions are built and the verification procedures that determine truth values. We have argued that semantic representations can be transparently mapped into verification procedures. When two equivalent semantic representations are being compared, as with the truth-conditionally equivalent (12a) and (12b), repeated here, examining the psychological processes implied by directly implementing these meanings as verification procedures can provide decisive evidence for distinguishing them.

(12) a. $|\text{DOT} \ \& \ \text{BLUE}| > |\text{DOT} \ \& \ \text{-BLUE}|$

b. $|\text{DOT} \ \& \ \text{BLUE}| > |\text{DOT}| - |\text{DOT} \ \& \ \text{BLUE}|$

Although these alternatives describe the same truth conditions, the psychological mechanisms required to implement them transparently are quite distinct. Whereas (12b) can be computed across all possible dot-flashing contexts using only the information provided directly by the visual system in concert with the approximate number system, (12a) is more psychologically brittle. Because it asks for information that cannot be directly provided by the visual system, it requires a context-driven transformation, identifying the set(s) in the context that are appropriate redescriptions of {DOT & -BLUE}. While such a transformation allows for accurate verification in contexts containing less than three colors of dots, this transformation would be less effective in contexts containing more than three colors of dots. However, as we have seen, the number of colors of dots played no role in explaining participants' *most* judgements, casting doubt on the hypothesis that they use a verification procedure based on (12a) in any context.

What may be surprising to consider, however, is that, in the context of only blue dots and dots of one other color (i.e., a 2-Color trial), the expression in (12a) would lead to more accurate performance in evaluating *most* than the expression in (12b). Specifically, with only two colors present in the array, (12a) is a more accurate verification procedure within the ANS than (12b).

This last point requires elaboration. Various studies have demonstrated that adults can rely on the Approximate Number System to make ordinal judgements (more/less) between two sets whether they are presented serially or in parallel (Dehaene 1997). In all cases, the estimated Weber Fraction for adults is better than the 3:4 value we found here for all trial types. Typically, the Weber Fraction for adults is closer to 7:8 (Dehaene 1997) and may be as high as 9:10 (Halberda & Feigenson, 2007; Piazza et al, 2003), and children as young as 4-years-old have a Weber Fraction of at least 3:4 (Halberda & Feigenson, 2007). For this reason, if participants had simply selected the set serving as the first argument of the > relation (e.g., the blue dots), and the set serving as the second argument (e.g., yellow dots) directly on a 2-Color trial and compared these using the ANS, as previous work has demonstrated they can, we would have observed a critical Weber Fraction of at least 7:8. That participants' performance is far below this suggests that they are relying on a representation of *most* like the expression in (12b), even when there are more accurate, truth-conditionally equivalent, methods of verification available (i.e., 12a)

This last observation provides the strongest evidence for the Interface Transparency Thesis. Even when there is a more precise algorithm that is native to the interface system, semantic judgments are driven by algorithms that transparently compute the relation expressed in the meaning. The semantic representation of *most* (i.e., its canonical specification) thus plays a determinative role in identifying the verification procedure for a sentence containing that word, at least when a transparent verification procedure is available. The fact that participants never employ the verification procedure most naturally associated with the canonical specification in (12a), even when that verification procedure is positively invited by the context and would yield the most accurate estimate of the truth of the expression provides compelling evidence against (12a) being the meaning of the expression.[7]

Of course, the fact that the verification procedure reflects precisely the structure of the meaning has a natural explanation in the set of circumstances in which *most* applies. The second argument of the > relation is not guaranteed by the world to have only one easily selectable property (e.g. yellowness), and, because of the 3-set limit on parallel enumeration (Halberda et al, 2006), the limitations of the psychological machinery would lead to drastically reduced

[7] Other results suggest that the present 2-Color results are not due to subjects "sticking with" a verification procedure that will work for every trial type (e.g. 2-5 Colors). Even when *only* 2-Color trials are presented adult performance is consistent with the meaning expressed in 3b and not with 3c (Pietroski et al, 2008).

performance as the heterogeneity of the remainder set increased. Thus, the most general verification procedure would be one that can apply independent of whether such a property exists in a particular circumstance. A verification procedure that varied as a function of contingent properties of the world would be less reliable than one which could apply across all circumstances.

Finally, we wish to reiterate that treating semantic hypotheses as psychological hypotheses makes available certain kinds of evidence that are unavailable to semantic theories concerned only with compositionally determined truth conditions, and moreover, that such evidence enables us to distinguish otherwise equivalent hypotheses. We have argued that semantic hypotheses are best viewed as psychological hypotheses about the mental representations involved in defining the truth conditions for a sentence. These representations provide canonical specifications of meaning that can be mapped transparently to verification procedures involving the integration of linguistic information with information from adjacent cognitive systems. Knowing what information these systems can and cannot provide places constraints on the verification procedures. And these constraints can, in turn, be used to examine the semantic representations themselves, enabling us to distinguish semantic hypotheses that are otherwise equivalent. We believe that this approach has so far been fruitful for distinguishing hypotheses about the meaning of *most*, but we view the demonstration that such questions can be precisely asked and plausibly answered as the more significant contribution of this work, opening the door for progress in the field of psycho-semantics.

References

Barwise, J. and R. Cooper (1981) Generalized Quantifiers in Natural Language. *Linguistics and Philosophy*, 4,159-219.

Brannon, E. M., Lutz, D. & Cordes, S. (2006). The development of area discrimination and its implications for number representation in infancy. *Developmental Science* 9 (6), F59–F64.

Chomsky, N. (1986) *Knowledge of Language: Its nature, origins and use*. Praeger: NY.

Church, Alonzo (1941) *The Calculi of Lambda Conversion*. Princeton University Press: Princeton, NJ.

Davidson, D. (1967) The logical form of action sentences. In *Essays on Actions and Events*. Clarendon Press: Oxford. 105-148.

Davies (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.

Dummett, M. (1973) *Frege, Philosophy of Language*. Cambridge: Harvard University Press.

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 Science*, 8, 307-314.

Hackl, M. (2008) On the Grammar and Processing of Proportional Quantifiers: Most versus More Than Half. *Natural Language Semantics*, Article in press.

Halberda, J. & Feigenson, L. (2008). Developmental change in the acuity of the "Number Sense": The approximate number system in 3-, 4-, 5-, 6-year-olds and adults. *Developmental Psychology*, 44(5), 1457-1465.

Halberda, J., Mazzocco, M.M.M. & Feigenson, L. (2008). Differences in primitive math intuitions predict math achievement. *Nature*, 455, 665-668.

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

Halle, M. (2002) *From memory to speech and back*. Mouton de Gruyter: The Hague.

Higginbotham, J. & R. May (1981) Questions, Quantifiers & Crossing. *The Linguistic Review*, 1, 41-80.

Horty, J. (2007) *Frege on definitions*. Oxford University Press: Oxford.

Jacobson, R., G. Fant & M. Halle (1952) *Preliminaries to speech analysis: The distinctive features and their correlates*. Technical Report 13. Massachusetts: Acoustics Laboratory, MIT.

Jackendoff, R. (1983) *Semantics & Cognition*. MIT Press: Cambridge.

Jackendoff, R. (1990) *Semantic Structures*. MIT Press: Cambridge.

Jackendoff, R. (2002) *Foundations of Language*. Oxford University Press: Oxford.

Jusczyk, P. (1997) *The discovery of spoken language*. MIT Press: Cambridge.

Kuhl, P. (1993) Early linguistic experience and phonetic perception: Implications for theories of developmental speech perception. *Journal of Phonetics*, 21, 125-139.

Landau, B. & R. Jackendoff (1993) "What" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences*, 16(2), 217-238.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & M. Studdert-Kennedy. (1967). Perception of the speech code. *Psychological Review*, 74, 431-461

Liberman, A. M., & I. G. Mattingly. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1-36.

Marr (1982) *Vision*. MIT Press: Cambridge.

Montague (1970) Universal grammar. *Theoria,* 36, 373–398.

Mostkowski, A. 1957. On a Generalization of Quantifiers. *Fundamenta Mathematicae*, 44, 12-36.

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., J. Lidz, T. Hunter & J. Halberda. (2008). The meaning of most: Semantics, numerosity and psychology. *Mind & Language*. Article in press.

Poeppel, D., Idsardi, W. van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society of London* B 363:1071-86.

Stevens, K. (1972) "The quantal nature of speech: evidence from articulatory-acoustic data." In E.E. David, Jr. & P.B. Denes, éds., *Human Communication: a Unified View*. New York: McGraw-Hill, pp. 51-56

Tarski, A. (1944). "The Semantical Concept of Truth and the Foundations of Semantics," *Philosophy and Phenomenological Research* 4: 341-75.

Treisman, A. & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95(1), 15-48.

Treisman, A. & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: Human Perception & Performance*, 16(3), 459-478.

Werker, J.F. (1995) Exploring developmental changes in cross-language speech perception. In D. Osherson (series editor), *An invitation to cognitive science*. L. Gleitman & M. Liberman (Volume editors), *Part I: Language*. Cambridge, MA: MIT Press. (pp. 87-106)

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

Wolfe, J.M. (1998). Visual search. In Pashler, H. (Ed.), *Attention*. Hove, England: Psychology Press/Erlbaum, UK Taylor & Francis, 13-73.

Acknowledgements

J.H. and J.L. devised the task; J.H., J.L., P.P., and T.H. defined the trial types of interest; T.H. implemented and ran the experiment; J.H. analyzed the data; J.L. wrote the manuscript with input from J.H., P.P. and T.H..

Corresponding Author

Correspondence and requests for materials should be addressed to J.L. (jlidz@umd.edu).

Rights of subjects

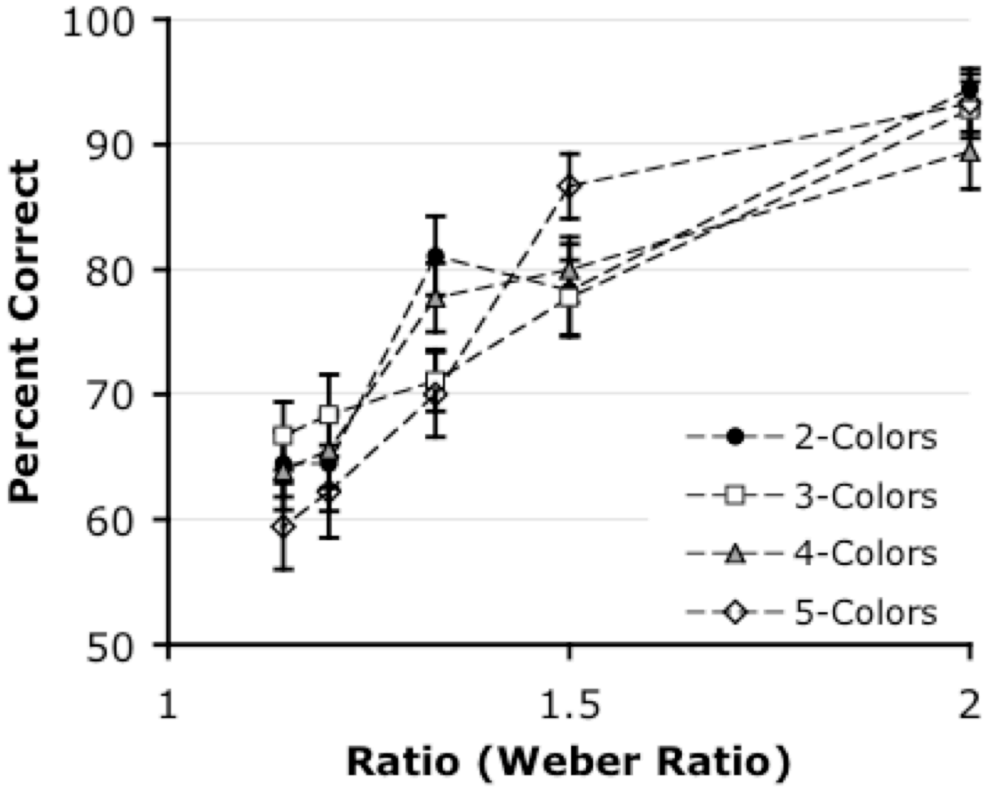
Guidelines for testing human research subjects were followed as certified by the Johns Hopkins University and The University of Maryland Institutional Review Boards. Subjects' rights were protected throughout.

Table 1. Parameter estimates from psychophysical model

Trial Type	R^2	Critical Weber Fraction	Nearest Whole-Number Ratio
2-Colors	.9480	.290	3:4
3-Colors	.9586	.320	3:4
4-Colors	.9813	.283	3:4
5-Colors	.9625	.316	3:4

Figure Captions

Figure 1. asdfjkl



100
90
80
70
60
50
Percent Correct
1
1.5
2
Ratio (Weber Ratio)
2-Colors
3-Colors
4-Colors
5-Colors

Figure 2

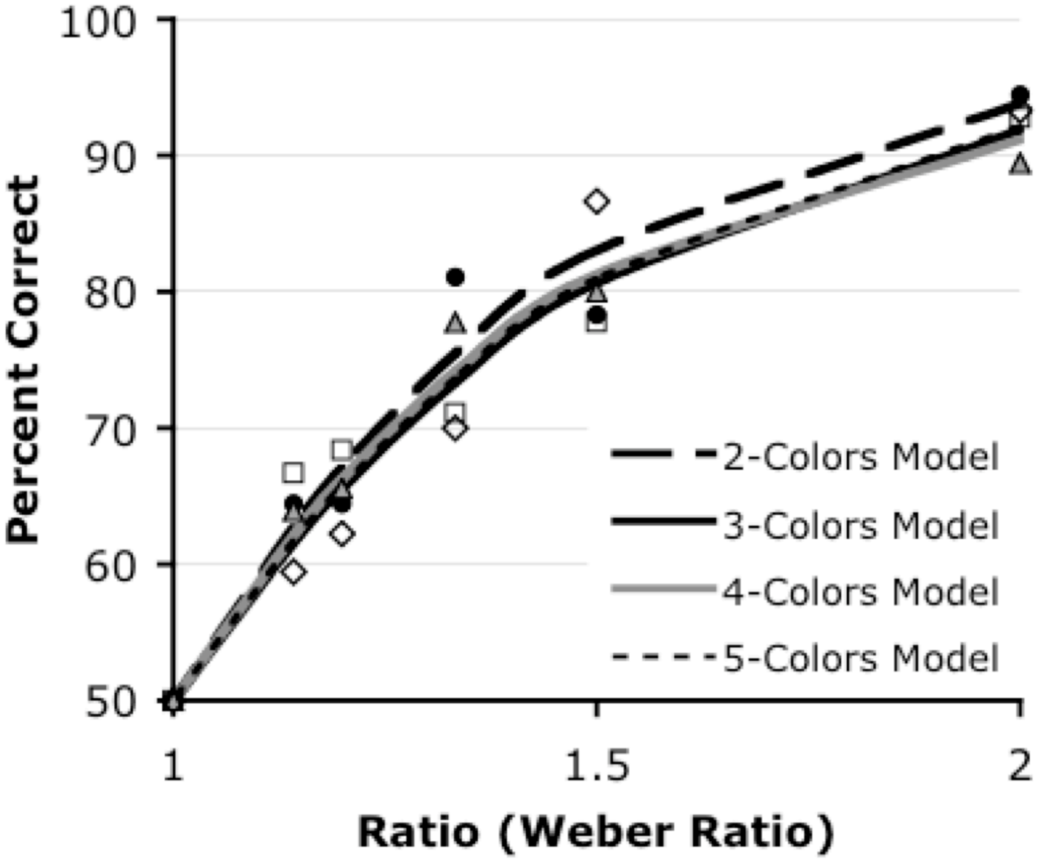

100
90
80
70
60
50
Percent Correct
1
1.5
2
Ratio (Weber Ratio)
2-Colors Model
3-Colors Model
4-Colors Model
5-Colors Model

Figure 3

Notes