# Seeing What You Mean, Mostly

Draft #6

16-July-11

Paul Pietroski, University of Maryland, College Park
Jeff Lidz, University of Maryland, College Park
Tim Hunter, Yale University
Darko Odic, Johns Hopkins University
Justin Halberda, Johns Hopkins University

**Abstract:**

Idealizing, a speaker endorses or rejects a (declarative) sentence $\Sigma$ in a situation **s** based on how she understands $\Sigma$ and represents **s**. But relatively little is known about how speakers represent situations. Linguists can construct and test initial models of semantic competence, by supposing that sentences have representation-neutral truth conditions, which speakers represent *somehow*; cp. Marr's (1981) Level One description of a function computed, as opposed to a Level Two description of an algorithm that computes outputs given inputs. But this leaves interesting questions unsettled. One would like to find cases in which $\Sigma$ can be held fixed, while modifying **s** in ways that have predictable effects on the nonlinguistic cognitive systems recruited to evaluate $\Sigma$. Extant work in perceptual psychology offers opportunities for eliciting judgments from speakers in highly controlled settings where something is known about the cognitive systems that speakers recruit when endorsing or rejecting a target sentence. In such settings, behavioral data can reveal aspects of how the human language system interfaces with other systems of cognition that are presumably shared with other species. As an illustration, we focus on the quantificational word 'most' and how perception of numerosity is related to the meaning of 'Most of the dots are blue', in the hope that studies of other perceptual systems may provide analogous opportunities for investigating how words are related to prelinguistic representations.

What are lexical meanings? And how are they related to the mental representations that speakers use to evaluate declarative sentences for truth/falsity in contexts? We address these questions by means of a case study, using experimental methods to reveal how the English word 'most' is related to perception of numerosity. Much of our discussion is restricted to examples like (1),

(1) Most (of the) dots are yellow

with 'most' as a quantificational determiner that combines with a plural noun or a plural partitive phrase; see also Hackl (2009). But we'll end with some speculations about adjectival 'most' and mass nouns, as in (2-4), and the relevance of our case study to a wider class of constructions.[1]

(2) I have the most dots

(3) Most of the paint is yellow

(4) Most of the dot is yellow

Following Chomsky (1995) and many others, we assume that humans have a faculty that generates linguistic expressions that interface with "perceptual-articulatory" and "conceptual-intentional" systems. Focusing on the latter interface, and stressing understanding as opposed to production, expressions can be described as instructions for how to build mental representations that exhibit "logical forms." If one dislikes talk of instructions, one can speak of blueprints for— or programs whose execution leads to the assembly of—certain representations. Or in honor of Frege's (1879) place in the history of semantics, one might introduce the technical notion '*Begriffsplan*', suggesting that the meaning of a generable expression is a "concept-plan" that can be paired with a "pronunciation-plan." But whatever the terminology, our questions here concern the structure and constituents of the representations whose assembly is guided by (1).

---

[1] We discuss other joint work that employs methods from the study of vision, suppressing many details reported in: Hunter et al. (2009), Pietroski et al. (2009), Lidz, et al. (in press), Halberda et al. (submitted), Odic et al. (submitted). And we rely on independently confirmed claims about perception of numerosity, in humans and other animals; for reviews, see Feigenson, Dehaene & Spelke (2004), Dehaene (1997), and the appendix to Lidz et al. (in press).

In our view, (1) is an instruction for how to build a thought that includes *arithmetic* concepts: #[DOT(X) & YELLOW(X)] > #[DOT(X)] – #[DOT(X) & YELLOW(X)]; the *number* of yellow dots *exceeds* the result of *subtracting* that number from the number of dots. Ignoring (2-4) for now, we think that 'most' accesses a complex quantificational concept: #[Φ(X) & Ψ(X)] > #[Φ(X)] – #[Φ(X) & Ψ(X)]; where 'Φ' and 'Ψ' correspond to the internal and external (restricting/scope) arguments of the determiner. But one might posit other underlying concepts, as in: #[DOT(X) & YELLOW(X)] > #[DOT(X) & ~YELLOW(X)], with a symbol for negation and not subtraction; #[DOT(X) & YELLOW(X)] > #[DOT(X)]/2, with division (or halving) instead of negation; the anumeric thought ONETOONEPLUS[DOT(X) & YELLOW(X), DOT(X) & ~YELLOW(X)], according to which some but not all of the yellow dots correspond one to one with all of the nonyellow dots. (In general, ONETOONEPLUS[Φ(X), Ψ(X)] iff there are more Φs than Ψs.)

Hackl (2009) suggests another option involving quantification: the plurality of yellow dots outnumbers *every other nonoverlapping* plurality of dots. And while this kind of proposal could be combined with a 'OneToOnePlus' gloss of 'outnumbers', Hackl assumes cardinality comparisons. Or perhaps 'most' does not call for any structured representation, and the logical form of (1) is simply MOST:DOT(X)[YELLOW(X)], because understanding a word is a matter of knowing that it has a certain satisfaction condition that can be specified in many ways.

In short, there are many ways in which one *might* understand 'most'. But experimental evidence (Hackl, 2009; Hunter et al., 2009; Pietroski et al., 2009; Lidz, et al., in press; Halberda et al., submitted; Odic et al., submitted) confirms (a) the ancient suspicion that a word of the "logical" vocabulary can indicate a complex concept, and in our view, (b) the specific proposal that 'most' invokes concepts of cardinality and subtraction. As discussed below, any version of (a) raises questions about how meaning is related to verification, and (b) raises questions about

3

why expressions interface with cognition in this way. Our proposal thus bears on debates concerning choices between "extensionally equivalent" semantic theories.[3] But the point is not merely that experiments can help adjudicate these debates.

We also want to stress that extant work in perceptual psychology offers opportunities for eliciting judgments from speakers in highly controlled settings where something is known about the cognitive systems that speakers recruit when endorsing or rejecting a target sentence. In such settings, behavioral data can reveal aspects of how the human language system interfaces with other systems of cognition that are presumably shared with other species. Idealizing, a speaker endorses or rejects a sentence $\Sigma$ in a situation $\mathbf{s}$ based on how she understands $\Sigma$ and represents $\mathbf{s}$. So one would like to find cases in which experimenters can perform two kinds of manipulations: hold $\Sigma$ fixed, while modifying $\mathbf{s}$ in ways that have predictable effects on the nonlinguistic cognitive systems recruited to respond 'yes' or 'no'; or hold $\mathbf{s}$ fixed, while modifying $\Sigma$.

Our illustration, based on English constructions and perception of numerosity, is offered in the hope that studies of other perceptual systems will provide further opportunities to investigate how words are related to prelinguistic representations. Linguists can still construct and test useful initial models of semantic competence by supposing that sentences have representation-neutral truth conditions, which speakers represent *somehow*; cp. Marr's (1981) Level One descriptions of functions computed, by the visual system and its various subsystems, as opposed to Level Two descriptions of algorithms that compute outputs given inputs. But if the goal is to study the human language faculty and its relation to other cognitive systems, theorists need sources of evidence for more psychologically detailed models of semantic competence.

---

[3] See Quine (1964), Davidson (1974), Foster (1976), Evans (1981), Peacocke (1986), Davies (1987). Hackl (2009) focused on 'most', in the specific context of applying Generalized Quantifier theory (Mostowski [1957]) to natural language semantics; see Barwise and Cooper (1981), Higginbotham and May (1981), Keenan and Stavi (1986).

Drawing on what is already known about human visual perception and numerical cognition, we offer empirical reasons for moving beyond representation-neutral specifications of truth conditions and towards specifying meanings in terms of specific algorithms (cp. Marr's Level Two) for assembling thoughts from available cognitive resources.

**1 'Meaning', Means, Meaning**

Pretheoretic considerations pull the word 'mean' in different directions, creating opportunities for terminological confusion, especially with regard to how meaning is related to analysis and verification. In this section, we make our mentalistic perspective explicit, highlighting the empirical issues addressed below. In section two, we defend our specific analysis of 'most', based on evidence concerning the evaluation procedures that speakers use in controlled contexts. The idea is that if (1) is used to assemble a thought of the form #[DOT(X) & YELLOW(X)] > #[DOT(X)] – #[DOT(X) & YELLOW(X)], then other things equal, speakers who can do so will evaluate (1) for truth/falsity by comparing the number of yellow dots to the result of subtracting that number from the number of dots. Evidence that speakers are biased towards this procedure, instead of available alternatives that would be as good or better in the contexts at hand, is prima facie evidence that (1) is used to assemble the posited thought. But we reject any form of verificationism according to which meanings are procedures for using experience to verify sentences.[4] Our view is not that 'most' is analyzable in terms of "sensory" concepts and linguistic "conventions". As discussed below, our claim is simply that data concerning how speakers verify sentences like (1) can provide evidence for how speakers understand 'most'.

*1.1 Meanings: Representations or Represented?*

---

[4] It is hard to see how such a view could accommodate the diversity and composability of word meanings. Likewise, we do not say that concepts like #[Φ(X) & Ψ(X)] > #[Φ(X)] – #[Φ(X) & Ψ(X)] are epistemologically privileged, compared with concepts like ONETOONEPLUS[Φ(X) & Ψ(X), Φ(X) & ~Ψ(X)]. For us, the question is how speakers understand 'most', not how 'most' is related to knowledge rooted in experience.

On the one hand, it can seem obvious that word meanings are mental representations of some kind, and hence that word meanings exhibit representational forms that cognitive scientists can try to characterize. Nonsynonymous words can be correlated with the same aspects of reality; consider 'Hesperus' and 'Phosphorus', 'woodchuck' and 'groundhog'. A meaningful expression, like 'Vulcan' or 'unicorn', need not correspond to any aspect of mind-independent reality. And polysemy suggests that a word (e.g., 'book' or 'France') need not correspond to a single aspect of reality. But on the other hand, speakers represent the world differently, yet still communicate. This might lead one to conclude that meanings are aspects of the environment that diverse speakers "triangulate on" when using language to convey information. From this externalist perspective, meanings do not exhibit representational forms.

We follow the former and older tradition of taking meanings to be representations of some kind. But we recognize the attractions of the following hypothesis: understanding the words of a natural spoken/signed language (i.e., knowing what the expressions mean) just *is* a matter of somehow assigning the right satisfaction conditions to the sounds/signs in question. Tarski (1933) showed how to recursively characterize satisfaction conditions for expressions of certain invented languages, and then characterize truth conditions for the boundlessly many sentences of such a language in terms of satisfaction. Davidson (1967) and Montague (1974) later outlined a program of specifying Tarski-style theories of truth for natural languages like spoken English. To take a trivial example, perhaps in French, 'rouge' is satisfied by an entity e— or more precisely, for any variable $v$, 'rouge($v$)' is satisfied by a sequence of entities that assigns e to $v$—if and only if e is red. And efforts to extend this progam have led to many insights.

Of course, one wants to know how a natural word comes to have a satisfaction condition, and how satisfaction conditions are related to the psychological phenomenon of understanding.

But in retrospect, an obvious suggestion is that competent speakers of French represent 'rouge' *as* a word satisfied by e iff e is red, and that speakers deploy this "tacit knowledge" in understanding expressions like 'baton rouge', which is known to be satisfied by e iff e satisfies both 'baton' and 'rouge'.[5] And one can speculate that speakers represent 'rouge' as having a certain satisfaction condition without any representation being an aspect of what 'rouge' *means*.

For each speaker, 'rouge' might call a specific concept to mind. But perhaps in part for this reason, each speaker understands the word in terms of some concept's *extension*, allowing for conceptual variation across speakers (and within a speaker) of the same language. Two individuals might link 'rouge' to different concepts, because one individual is color-blind, yet specify the same satisfiers. The concepts might be extensionally equivalent, at least if this kind of equivalence is sensitive to patterns of communication and deference. So especially if conceptual variation across speakers is the norm, one might think that extensional equivalence— across *all* expressions of a natural language—is good enough for sameness of understanding, and thus sameness of meaning. And perhaps even with regard to words like 'most', mature speakers of a language represent shared satisfaction conditions in different ways.

On this view, shared meanings are represented but representation-neutral. This conception of meaning/understanding is often combined with skepticism about appeals to any substantive analytic-synthetic distinction or semantic decomposition; see, e.g., Fodor and Lepore

---

[5] Davidson and Montague did not say that competent speakers tacitly know the axioms of a truth theory. At least for Davidson, the idea was simply that such knowledge would *suffice* for understanding; actual speakers may not satisfy this sufficient condition. By contrast, many subsequent advocates of truth conditional semantics offered an explicitly psychological hypothesis, often in connection with Chomsky's (1965, 1986) developing conception of grammar and the I-language/E-language distinction; see, e.g., Evans (1981), Higginbotham (1985), Peacocke (1986), Davies (1987), Larson and Segal (1995), Heim and Kratzer (1998). For skepticism about this psychologized version of the Davidson-Montague program, see Chomsky (2000), Pietroski (2005, 2010) and references there.

(2002). So the details of how speakers represent satisfaction conditions might seem to be semantically irrelevant. But we deny that meaning is representation-neutral in this way.

*1.2 Non-neutrality*

Suppose two speakers use different concepts—say, UNICORN(X) and GRYPHON(X)—to specify the satisfiers of a noun that has no satisfiers: one speaker represents the noun as being satisfied by e iff e is a unicorn; the other represents the noun as being satisfied by e iff e is a gryphon. Here, one suspects different lexical meanings (and in that sense, different words). One can claim that such examples should be set aside, as special cases. But this raises questions about whether 'groundhog' and 'woodchuck' differ semantically, or if apart from pronunciation, this is just another example of speakers representing a common extension in different ways. And in any case, if representational format is an aspect of meaning for "fictional vocabulary" items like 'gryphon', it may also be an aspect of meaning for "logical vocabulary" items like 'most'.

This would fit with the ancient observation that sentences containing logical vocabulary seem to exhibit *analytic* relations. Suppose there are a hundred marbles in a jar. Competent speakers know that (5) implies (6), which implies (7), but not conversely.

> (5) All of the marbles are red
>
> (6) Most of the marbles are red
>
> (7) Some of the marbles are red

This calls for explanation that is not provided simply by representing the logical forms as follows: ALL:MARBLES(X)[RED(X)]; MOST:MARBLES(X)[RED(X)]; SOME:MARBLES(X)[RED(X)].

A speaker who represents the semantic contribution of each determiner with an atomic concept may not represent the generalizations: ALL:$\Phi$(X)[$\Psi$(X)] $\supset$ MOST:$\Phi$(X)[$\Psi$(X)]; MOST:$\Phi$(X)[$\Psi$(X)] $\supset$ SOME:$\Phi$(X)[$\Psi$(X)]. By analogy, competent speakers may not know that if x

heats y, x raises the mean molecular energy of y. Perhaps for logical vocabulary, speakers do know various "meaning postulates." But another hypothesis is that the concepts expressed with quantificational determiners share parts. Perhaps (5) and (6) are understood as follows: #[MARBLE(X) & ~RED(X)] = 0; #[MARBLE(X) & RED(X)] > #[MARBLE(X) & ~RED(X)]. Or alternatively and anumerically: ONETOONE[MARBLE(X), MARBLE & RED(X)]; ONETOONEPLUS[MARBLE(X) & RED(X), MARBLE(X) & ~RED(X)].

Whatever the format, recognizing implications requires some nonsemantic knowledge (cp. Quine 1951), perhaps knowledge concerning cardinalities or correspondence/remainders (cp. Barner et al., 2010). The empirical questions concern the forms of the thoughts indicated with (5-7) and the extralinguistic resources required to recognize that these thoughts are *not* truth-conditionally independent. And while open-class vocabulary items (nouns, verbs, and adjectives/adverbs) typically resist analysis, logical vocabulary items (in particular, quantifiers and connectives) often invite analysis. It is notoriously hard to offer even one plausible decomposition of CHICKEN(X), which seems not to be a conjunction of ANIMAL(X) with any differentiating concept Φ(X); see, e.g., Fodor (1998, 2003). But with regard to 'most', the problem is that there are too many candidates—involving cardinality, OneToOnePlus, subtraction, negation, etc. In this respect, 'most' seems akin to *complex* expressions like 'chicken egg', which presumably *have* a compositional analysis even if the details elude theorists.[7]

These considerations have long motivated the study of quantificational expressions as a way to perhaps reveal core aspects of how linguistically expressible thought is structured; see Ludlow (2005). If (5-7) are understood as sentences that are not truth-conditionally independent,

---

[7] The concept CHICKEN-EGG invites at least three (extensionally similar) analyses: EGG-LAID-BY-A-CHICKEN, EGG-FROM-WHICH-A-CHICKEN-CAN-GROW, and EGG-THAT-IS-ALSO-CHICKEN (cp. 'yellow metal' and 'human child').

then for each of these sentences, formally distinct specifications of its truth condition invite corresponding proposals about the representational forms that speakers employ. If speakers represent the truth conditions in a common way, despite having the concepts required to exhibit representational diversity, that undermines motivation for an externalistic representation-neutral conception meaning. And if speakers uniformly understand 'most' in terms of cardinalities and subtraction, perhaps representational details are semantically relevant after all. Studies of particular lexical items can thus bear on questions about how meaning is related to cognition.

In the case of 'most', it is also relevant that linguistic understanding and cognitive development unfold in parallel. In particular, young children who have not yet attained numerical competence can exhibit at least partial understanding of 'most'. (Barner, Chow & Yang, 2009; Barner, Libenson, Cheung & Takasaki, 2009; Halberda, Taing & Lidz 2008; Halberda, Hunter, Pietroski & Lidz, submitted; Papafragou & Schwarz, 2005/6). More generally, one might wonder how a speaker who cannot count and reliably enumerate things *could* acquire and maintain *any* concept (or word) whose extension can be specified in terms of cardinalities— e.g., as $\#[\Phi(x) \& \Psi(x)] > \#[\Phi(x) \& \sim\Psi(x)]$. Some things can out*number* some other things, even if the former are "lesser" along many noncardinal dimensions, like area or mass. So in this sense, comparing cardinalities requires a capacity to abstract from various perceptible respects in which some things are "more" than others.

Nonetheless, humans and other animals can represent cardinalities in an approximating way—without counting—across a wide range of perceptible features often exhibited by countable things. This natural capacity is supported by an "Approximate Number System" (ANS) that is ratio dependent, in accord with Weber's Law, with accuracy/discriminability declining as number increases. Given a scene with 66 red and 34 blue marbles, it is easy to

"sense" that more are red. On the other hand, given 51 red and 49 blue, judgments will be almost at chance.[8] And speakers might use their ANS to evaluate sentences like (6) for truth/falsity.

(6) Most of the marbles are red

This can still seem to tell *against* analyzing MOST:Φ(x)[Ψ(x)] in terms of any concept extensionally equivalent to #[Φ(x) & Ψ(x)] > #[Φ(x) & ~Ψ(x)]. One might imagine that instead, children use their ANS to acquire a 'most'-prototype and then an atomic concept that is not *logically* related to other atomic concepts. Note that if 'most' accesses an atomic concept—i.e., if competent speakers represent the indicated satisfaction condition atomically—there is no guarantee that 'most' has the "simple majority" satisfaction condition specified in various ways above. One can imagine a quantificational word 'smost' such that 'Smost of the marbles are red' means roughly that a *supermajority* of the marbles are red, with speakers being sure that 51% is not enough, unsure about 60%, but sure that 70% is enough. And if 'most' is semantically atomic, its meaning might well be that of 'smost'.

Indeed, one might think that even if a community of numerate adults stipulated a simple majority meaning, generations of innumerate children would eventually acquire a word with the meaning of 'smost'. But as discussed in section two, our experiments suggest that 'most' is understood as a word that makes (6) true so long as the number of red marbles exceeds the number of other marbles—no matter how small the difference, even if counting is impossible.

---

[8] For reviews and relevant literature, see Dehaene (1997); Feigenson, Dehaene and Spelke, (2004); Feigenson (2007); Cantlon, Platt, and Brannon (2009); the appendix to Lidz. et.al. (in press), and references there. 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 ratio effects (Moyer & Landauer, 1967). And in numerate cultures, five-year-olds have mapped representations of the ANS onto the discrete number words (Le Corre & Carey, 2007). There have also been suggestions that the ANS plays some associated role in formal mathematics (Halberda, Mazzocco, & Feigenson, 2008) and math learning disabilities (Piazza et al., 2010). Crucially for present purposes, the ANS can generate cardinality estimates for up to three perceptual concepts in parallel, for both human adults (Halberda, Sires, & Feigenson, 2006) and infants (Zosh, Halberda & Feigenson, 2011). Recording of single neurons in monkeys suggests that the ANS can generate a representation of the approximate number of items present within 150 ms of stimulus onset (Nieder & Miller, 2004).

Given pragmatics, uttering (6) may still suggest that significantly more than half of the marbles are red. But if this is not what (6) *means*, one wants to know why. We suspect that children represent the semantic contribution of 'most' by choosing from a limited menu of concepts that can be accessed by the human language system to construct thoughts that get evaluated (e.g., for truth/falsity) via other cognitive systems. The idea is that given these constraints, the nonatomic concept $\#[\Phi(x) \mathbin{\&} \Psi(x)] > \#[\Phi(x)] - \#[\Phi(x) \mathbin{\&} \Psi(x)]$ is an especially good choice for 'most'.

Obviously, no single experiment or phenomenon will provide decisive evidence in favor of any specific hypothesis about representational format. But if animal minds make repeated use of certain circuitry that supports computations that are required across many cognitive domains (see, e.g., Gallistel, 1989) then one should not be surprised if logical vocabulary in a human language accesses such circuitry in a systematic way that gets manifested in many small ways. In section two, we focus on studies designed to reveal how adults manage to endorse or reject sentences like (6) in contexts where counting is not possible but other strategies are available. But many other facts can be relevant.

For example, Halberda et al. (submitted) report that a nontrivial percentage of children go through a phase of understanding 'most' as if it meant *fewer* than half. One can speculate that for such children, 'most' is semantically atomic: they need to correct a nonlinguistic/nonlogical error about when the adult word applies. But this leaves a puzzle: of all the possible errors, why this one? Another thought, requiring further study to confirm, is that the nonatomic concept $\#[\Phi(x) \mathbin{\&} \Psi(x)] < \#[\Phi(x)] - \#[\Phi(x) \mathbin{\&} \Psi(x)]$ is on a child's menu of options for determiner meanings. Perhaps 'most' always calls for comparing the cardinality of a *conjunction*, of a restrictor concept $\Phi(x)$ with a scope concept $\Psi(x)$, to a *subtraction* (of that cardinality from the restrictor's

12

cardinality).[9] Some children might take the comparison to be *less than*, instead of *greater than*, because this would contrast with the salient competitor 'more'.

In other work, we look for evidence that meaning is not representation neutral by focusing on respects in which (6) differs from (8), even when all the marbles are blue or red;

> (8)  More of the marbles are blue than red

see Halberda et al. (submitted). We mention these various sources of evidence, in part to stress that our aim is *not* to identify meanings with verification procedures. Our proposal is that 'most' is understood in a certain way, with testable implications for how speakers evaluate sentences. And one can reject implausible forms of verificationism without denying (the truism) that logical forms often invite certain ways of evaluating whether or not thoughts of that form are true.

## 1.3  Logical Form and Verification

Suppose that one is asked, in a context that makes the question reasonable, whether Al sang and Bo danced. In answering, one might execute the following procedure: find out whether or not Al sang; find out whether or not Bo danced; if the answer to each sub-question is affirmative, answer the posed question affirmatively, and otherwise answer negatively.

In principle, there are endlessly many other strategies that might be reasonable in particular contexts: asking a friend who was at the relevant party (and repeating her answer); checking the calendar and answering affirmatively if and only if last night was May 5th (since Al and Bo perform only and always on that date); etc. But it seems plausible that other things being equal, if a competent speaker of English can evaluate each of two thoughts for truth/falsity, then

---

[9] This ensures that 'most' is "conservative" in Barwise and Cooper (1981) sense. That is, DETERMINER:$\Phi(x)[\Psi(x)]$ iff DETERMINER:$\Phi(x)[\Phi(x) \& \Psi(x)]$; see also Higginbotham and May (1981), along with Keenan and Stavi (1986) on defining complex determiners. As children acquire 'most', their understanding can be non-adult in various ways; see Papafragou and Schwarz (2005/2006), Barner et al. (2009). Though one must consider the possibility that some children are simply ignoring a word they do not understand, effectively treating instances of 'Most $\Phi$s are $\Psi$s' as bare plural generics with an existential interpretation perhaps supplemented with stronger pragmatic implication.

she will do so if she needs to evaluate their conjunction for truth/falsity. Indeed, one might suspect that even if one asks a friend, one uses the friend's testimony that Al sang and Bo danced as evidence that Al sang *and* as evidence that Bo danced. In which case, one could still be using the default evaluation strategy suggested by the logical form of a conjunctive claim: evaluate each conjunct, and conjoin the answers. Perhaps there are some cases in which speakers evaluate conjunctive claims more holistically, without evaluating the conjuncts. But if so, these are plausibly exceptions to the default strategy, perhaps involving metalinguistic judgments.

More generally, analyses suggest evaluation strategies. This is a mundane point, given that speakers can evaluate claims of arbitrary complexity; see Dummett (1973), Horty (2007). Sentences of the form 'P and Q' indicate logically complex thoughts, whose complexity corresponds to a potential procedure for determining whether the thoughts are true or false. Likewise, one can decide whether something is a cow that smiled by deciding (i) whether it is a cow and (ii) whether it smiled. This is so, because 'cow that smiled' indicates a concept that can be analyzed as COW(X) & SMILED(X). Prima facie, the default strategy for evaluating (a thought assembled via) '~P' is to evaluate 'P' and give the opposite answer. Again, there may be special cases in which evaluating 'P' is hard, and thinkers can employ strategies for evaluating '~P' more holistically. But if only because negation often creates difficulties for nonlinguistic systems, one suspects that other things equal, 'Al did not dance' will be evaluated (if possible) by deciding whether or not Al danced.

Turning from phrasal to lexical decomposition, one can at least imagine a language with a morpheme 'ne' in which (9) has the following logical form: ~(AL SANG) & ~(BILL DANCED),

(9) Al sang ne Bill danced

14

as opposed to (AL SANG) * (BILL DANCED); where '*' stands for an atomic connective such that (P) * (Q) iff neither P nor Q . In principle, theorists might discover this fact about (9) in many ways. But suppose experimental results suggested that other things equal, speakers evaluate (9) by evaluating each clause and conjoining the negations of the evaluations. That would, we claim, be prima facie evidence for the corresponding analysis of 'ne'—as opposed to positing an atomic concept with the same truth table. And in our view, the same points apply to 'most'. If evidence suggests that speakers are biased towards evaluating (1)

(1) Most of the dots are yellow

by comparing (their estimate of) the number of yellow dots to the result of subtracting that number from (their estimate of) the number of dots, even when other strategies are available, that is prima facie evidence for the corresponding analysis of 'most'.

To be sure, it can be hard to know if a certain evaluation procedure (deployed in certain contexts) is really a default procedure corresponding to the logical form in question, as opposed to a method of confirmation that happens to be favored by certain contexts. But that is a general point, which applies equally to any proposed analysis of any expression. One should not assume that evidence concerning how speakers evaluate claims cannot confirm proposed analyses. Especially if speakers fail to use available evaluation procedures that are favored by the context, such evidence can help adjudicate between alternative conceptions of logical form.

None of this implies or even suggests that meanings are generic procedures for using experience to confirm or disconfirm sentences. We readily grant that confirmation can be holistic and open-ended in ways that semantic composition is not. We think of meanings as procedures for building thoughts, with lexical meanings as atomic procedures for "fetching" concepts that

15

can be combined with others.[10] But the *structure* of an assembled thought can surely make some ways of evaluating the thought—gather certain information, by whatever means, and perform certain operations if you can—more natural than others. Taking this as a starting point, Lidz et al. (in press) urge an "Interface Transparency Thesis:" the verification procedures that speakers employ, when evaluating sentences they understand, are biased towards algorithms that directly compute the relations and operations encoded by the relevant logical forms. In §2.4, we return to and expand on this claim.

From this perspective, the real issues concern the kind of concept fetched with 'most'. Is this word representationally neutral, in that it can be used to access *any* concept mathematically equivalent to $\#[\Phi(x) \& \Psi(x)] > \#[\Phi(x)] - \#[\Phi(x) \& \Psi(x)]$? And if not, is there evidence that 'most' is used to access a concept with which one thinks about cardinalities and subtraction?

## 2 Computing 'Most'

In section one, we mentioned various possibilities for the logical form of (1).

(1)  Most (of the) dots are yellow

These options can be organized in terms of the atomic concepts posited, thereby suggesting a research strategy for gathering evidence for or against specific proposals.

*2.1  Possible Algorithms*

One option is to eschew decomposition and posit an atomic correlate of 'most', as in (1a).

(1a)  MOST:DOT(X)[YELLOW(X)]

As noted above, this proposal faces its own difficulties, and it has no special status as the null hypothesis. Moreover, any evidence that 'most' fetches a specific complex concept will be

---

[10] On this view (see Pietroski, 2010), the meaning of 'loves Juliet' (or 'most dots') can be described as an instruction to *fetch* two concepts from lexical "addresses" and *saturate* one with the other: **SATURATE[FETCH@'LOVES', FETCH@'JULIET']**; where executing this mini-program can yield the specific concept LOVES(JULIET).

evidence that the concept fetched is not atomic. And while we do not here assume that 'most' is *morphologically* complex (e.g., 'many/much-est'), arguments for such complexity—see e.g., Bresnan (1973), Bhatt and Pancheva (2004), Hackl (2009) drawing on Heim (1999)—bolster the old suspicion, articulable in terms of Frege's (1879, 1884) second-order logic, that many entailments of (1a) reflect complexity of the logical concept MOST:$\Phi(X)[\Psi(X)]$. So henceforth, we'll focus on versions of the idea that 'most' is not semantically atomic. And in terms of organizing hypotheses, one major question is whether 'most' is understood in terms of *cardinalities*; where cardinalities can serve as abstract "measures" that group concepts into equivalence classes, in terms of how many things fall under them.[11]

Various considerations initially suggest a negative answer. Studies of object-tracking in infants (Feigenson & Carey, 2003; Wynn, 1992) have revealed a system that can detect one-to-one correspondences, at least in certain situations. But as noted above, one might wonder how speakers who cannot count could understand 'most' in terms of cardinalities. And given how cardinality is related to correspondence, truth conditions specified in terms of cardinalities can often be specified in anumeric terms. In particular, (1b)

(1b)  ONETOONEPLUS[DOT(X) & YELLOW(X), DOT(X) & ~YELLOW(X)]

is true iff *some* (but not all) of the yellow dots correspond one to one with (all of) the nonyellow dots. So like (1), (1b) is true iff each nonyellow dot can be paired with its own yellow dot, leaving some yellow dots as "remainders" that are not paired with any nonyellow dot. By contrast, (1c) specifies this truth condition in terms of a comparison between numbers.

(1c)  #[DOT(X) & YELLOW(X)] > #[DOT(X) & ~YELLOW(X)]

---

[11] Restricting attention to concepts that have extensions, the *number of* a concept is the number of things falling under it; see Frege (1884). Deferring discussion of concepts like MUD(X), the number of $\Phi(x)$ is also the cardinality of $\{x: \Phi(x)\}$. So concepts are numerically equivalent iff their extensions have the same cardinality.

Recall that '#' stands for a mental symbol that combines with a monadic concept to form a complex representation of a number (viz., the cardinality of the concept/extension). So one might think that for speakers who cannot count, (1b) is a more plausible logical form for (1) than (1c); see Pietroski et al. (2009) for discussion. Indeed, one might suspect that speakers who can represent numbers are somehow exploiting tacit knowledge of the generalization stated in (10).

(10)   $\#[\Phi(x)] = \#[\Psi(x)]$ iff ONETOONE$[\Phi(x), \Psi(x)]$

This principle lies at the heart of arithmetic.[13] And one might think that at least typically, tacit knowledge of this principle covaries with being a counter; cp Heck (2000).

On the other hand, if humans and other animals can *estimate* cardinalities—by means of the ANS system noted above—then (1c) is more plausible. But in any case, there are various ways of specifying the numbers to be compared. While (1c) employs negation to specify the dots that are not yellow, (1d) employs subtraction and a representation of the number of dots.[14]

(1d)  $\#[\text{DOT}(x) \& \text{YELLOW}(x)] > \#[\text{DOT}(x)] - \#[\text{DOT}(x) \& \text{YELLOW}(x)]$

And of course, there are other ways to avoid a negation operator, as in (1e).

---

[13] The cardinality of concept $\Phi(x)$ *is* the cardinality of concept $\Psi(x)$—these concepts are numerically equivalent (see note 11)—iff the things falling under $\Phi(x)$ correspond one to one with the things falling under $\Psi(x)$. And this generalization, from which the axioms of arithmetic can be derived (in a second-order logic) given Fregean definitions of notions like 'zero' and 'predecessor' (see Heck (1993), Demopolous (1994), Zalta (2003)), can be recast in a way that explicitly treats cardinalities as classes of concepts: for every concept $\Phi(x)$, there is a class of concepts such that $\Psi(x)$ is the class iff ONETOONE$[\Phi(x), \Psi(x)]$; see Boolos (1998).

[14] In (1d), '−' stands for (a concept of) *cardinality* subtraction, which differs from *set* subtraction. For present purposes, one could rewrite (1c) as (1c′);

(1c′)  $|\{x:\text{DOT}(x)\} \cap \{x:\text{YELLOW}(x)\}| > |\{x:\text{DOT}(x)\} - \{x: \text{YELLOW}(x)\}|$

where the vertical bars indicate a mapping from *sets* (extensions of concepts) to cardinal numbers, and '-' (like '∩') indicates a mapping from pairs of sets to sets. In principle, one can ask fine-grained questions about whether negation is represented as an operator on concepts themselves or in terms of their extensions and set-subtraction. But we won't, since in our view, the logical form of (1) is more like (1d) than (1c) or (1c′). One can also ask fine-grained questions about whether the numbers in question are represented as the numbers of concepts (cp. Frege (1884)), or as the cardinalities of sets—as in (1d′) or (1d″).

(1d′)  $|\{x:\text{DOT}(x) \& \text{YELLOW}(x)\}| > |\{x:\text{DOT}(x)\}| - |\{x:\text{DOT}(x) \& \text{YELLOW}(x)\}|$

(1d″)  $|\{x:\text{DOT}(x)| \cap \{x:\text{YELLOW}(x)\}| > |\{x:\text{DOT}(x)\}| - |\{x:\text{DOT}(x)| \cap \{x:\text{YELLOW}(x)\}|$

We set these questions aside; though in our view, they are important.

(1e)  #[DOT(X) & YELLOW(X)] > #[DOT(X)]/2

Hackl (2009) argues against this last analysis, and more generally, against analyzing

'most' as 'more than half'. Our results confirm his conclusion by providing evidence for a

logical form specified in terms of cardinality subtraction. We will address Hackl's own proposal:

the number of yellow dots exceeds that of every other nonoverlapping set of dots. But initially,

we focus on (1b), (1c), and (1d). These options illustrate a pair of major divisions with respect to

algorithms for answering the question posed by (1): algorithms that require representations of

*cardinalities* versus those that do not; and within the first class, algorithms that *compute* one

cardinality in terms of another (e.g., by subtracting one number from another) versus those that

do not (e.g., by representing the cardinality of a concept that has a negated constituent).

With regard to the first distinction, between (1b) and (1c)/(1d), the point is *not* that

comparison of cardinalities will not involve detection of one to one correspondence (and

remainders) at any level of representation. The point is rather that detection of one to one

correspondence does not require representation of cardinalities. A thinker can figure out that

there are more cows than horses in a certain pen by pairing each horse with exactly one cow, and

noticing that this pairing leaves at least one cow unpaired with any horse. Alternatively, one can

represent the number of cows and the number of horses, and then compare those representations.

Of course, one way of comparing cardinal numbers is to represent them in terms of the

successor function, so that the fourth and third positive natural numbers are represented as

follows:  $0''''$, $0'''$. Then one can use a OneToOnePlus procedure to determine if the former

*exceeds* the latter: keep removing a prime ('') from each representation in turn, starting with the

19

former; answer 'yes' iff the primes on the latter representation are exhausted first.[15] Nonetheless, one can distinguish two kinds of procedures for evaluating the thought that most of the animals in the pen are cows: OneToOnePlus procedures that do not involve numeric representations (or storage of such representations); and cardinality-comparison procedures that do involve such representations and *may* involve a OneToOnePlus procedure (e.g., at the comparison stage).

It might help to think about the second distinction, between (1c) and (1d), in terms of the contrast between directly *representing* the cardinality of the nonyellow dots and *computing* that cardinality without representing the nonyellow dots as such. Imagine two different minds. The first can represent the dots that are not yellow, and represent their number (by counting, estimating, consulting an oracle, or whatever), just as it can represent the yellow dots and their number; but it is unable to subtract, or unable to represent the number of nonyellow dots *as* the result of subtracting the number of yellow dots from the number of dots. The second mind can (subtract and) represent the number of nonyellow dots in this computationally complex way; but it cannot represent the number of nonyellow dots as such, except in special cases where metalinguistic representations and overt counting procedures (or oracles) are deployed.

Such distinctions can be viewed as examples of Marr's (1981) contrast between functions computed and algorithms used, even if the issues concern semantic competence and not real-time generation of representations in response to stimuli. Theorists can and did initially describe semantic competence from a truth-conditional perspective that treats (1b), (1c), and (1d) as notational variants. But given any abstract mapping from sounds to truth conditions, one can go on to ask *how* speakers map inputs to outputs in terms of suitably composable representations

---

[15] In any base $n$, a longer representation indicates a greater number; and for representations of the same length, the comparison task can be reduced to comparison of numbers less than $n$. Indeed, given the generalization in (10), one should be unsurprised if cardinality-comparison algorithms exploit one-to-one comparison at some point.

that are accessible via lexical items. This shifts perspective, from idealized representation-independent extensions to idealized programs/instructions for assembling representations of a certain sort. Chomsky (1986) likewise contrasts I-languages, procedures that generate expressions, with E-languages that can be identified with sets of generable expressions.[16]

We end this subsection with a digression, recognizing that any appeal to cardinality concepts raises a foundational question: what do thinkers represent with representations like #[DOT(X)]? One can restrict attention to "count concepts," and say that #[DOT(X)] represents a cardinal number. But that leaves the question of what cardinal numbers (zero, one, two, etc.) are. And while cognitive scientists need not be metaphysicians, they do need to specify the *contents* of posited representations: *what* do thinkers represent with #[DOT(X)]? At least since Frege and Russell, one answer has been that cardinal numbers—in contrast to noninteger ratios and irrational numbers like π—are equivalence classes of extensions, or things that have extensions (e.g., mappings from entities to truth values, or idealized representations that are satisfied by entities). Such equivalence classes can be ordered, added, and subtracted in obvious ways.[17] So at least relative to the idealization that count concepts like DOT(X) have extensions, perhaps #[DOT(X)] represents an equivalence class of such concepts.

---

[16] Both contrasts, via Marr and Chomsky, are descendants of Church's (1941) distinction between two interpretations for expressions of his lambda calculus: as devices for referring to functions "in extension" (sets of input-output pairs); or as devices for referring to functions "in intension" (procedures for computing outputs given inputs). Read the former way, $\lambda x.x+1 = \lambda x.(3x+27/9)/(5-2)$, since the same set is described twice; read the latter way, $\lambda x.x+1 \neq \lambda x.(3x+27/9)/(5-2)$, since the procedures differ. We focus on the construal of 'λ' according to which $\lambda\Phi.\lambda\Psi.\text{ONETOONEPLUS}[\Phi(X) \& \Psi(X), \Phi(X) \& \sim\Psi(X)]$ differs from $\lambda\Phi.\lambda\Psi.\#[\Phi(X) \& \Psi(X)] > \#[\Phi(X) \& \sim\Psi(X)]$, which differs again from $\lambda\Phi.\lambda\Psi.\#[\Phi(X) \& \Psi(X)] > \#[\Phi(X)] - \#[\Phi(X) \& \Psi(X)]$.

[17] For any equivalence classes C and C′, C > C′ iff for any concepts $\Phi(x)$ in C and $\Psi(x)$ in C′, ONETOONEPLUS[$\Phi(x), \Psi(x)$]. Correlatively, C + C′ is the equivalence class C″ such that: for any concepts $\Phi(x)$ in C and $\Psi(x)$ in C′ such that $\sim\exists x[\Phi(x) \& \Psi(x)]$, the disjunctive concept [$\Phi(x) \text{ v } \Psi(x)$] is in C″; where (ideal) concepts are productively systematic in ways that ensure that any distinct equivalence classes will contain nonoverlapping concepts. Likewise, if C > C′, then C – C′ is the equivalence class C″ such that: for any concepts $\Phi(x)$ in C and $\Psi(x)$ in C′ such that $\forall x:\Psi(x)[\Phi(x)]$, the conjunctive concept [$\Phi(x) \& \sim\Psi(x)$] is in C″.

From this perspective, if an animal has (i) a concept DOT(X) with which it can think about dots as such, and (ii) a higher-order concept #[Φ(X)] that can be saturated by a concept like DOT(X), then the number represented with #[DOT(X)] is the equivalence class C such that for any concept Φ(X): Φ(X) is in C iff ONETOONE[Φ(X), DOT(X)]. Putting the point in an explicitly Fregean way, representations of numbers (as objects) are derivative on higher-order representations akin to FOUR[Φ(X)], which can be saturated by a concept like DOT(X) to yield a truth-evaluable thought.[18] We'll return to some related issues in section three. But having flagged the foundational question, and noted that answers need not be mysterious, we henceforth assume that cognitive scientists may legitimately appeal to representations of numbers.

## 2.2 Against OneToOnePlus

To provide evidence against hypothesis (1b), and in favor of an alternative like (1c) or (1d),

(1b) ONETOONEPLUS[DOT(X) & YELLOW(X), DOT(X) & ~YELLOW(X)]

(1c) #[DOT(X) & YELLOW(X)] > #[DOT(X) & ~YELLOW(X)]

(1d) #[DOT(X) & YELLOW(X)] > #[DOT(X)] − #[DOT(X) & YELLOW(X)]

one wants an experimental context with several features. First, speakers should be able to evaluate (1) in ways significantly better than guessing in the context.

(1) Most of the dots are yellow

Second, the stimuli in the context should vary with regard to ease of using a OneToOnePlus algorithm that can be used successfully in the experimental task, at least in the easy cases. Third,

---

[18] The higher-order concept FOUR[Φ(X)] clearly determines an equivalence class of concepts—viz., those that apply to four things. But the plural/nondistributive concept FOUR(X), which applies to some things iff they are four, is also of interest. Given some objects identified with first four cardinal numbers—e.g., ∅, {∅}, {{∅}}, and {{{∅}}}—they are four. And if Φ(X) is a concept that applies to each of these four special entities and no others, then for any other concept Ψ(X): FOUR[Ψ(X)] iff ONETOONE[Φ(X), Ψ(X)]; and FOUR[Ψ(X)] iff some things that collectively fall under FOUR(X) are such that they are all and only the things that fall (distributively) under Ψ(X),

presentation times should be long enough to engage speakers' default mechanisms of evaluation, but brief enough to discourage use of other strategies.

When display times are long, speakers can use top-down strategies that sometimes improve and sometimes degrade performance. For example, when given 800 milliseconds to estimate the number of dots on a screen, people make *less* accurate estimates for higher numbers (e.g, 30) compared with estimates made in 200 milliseconds; see Mandler & Shebo (1982). Attempts to engage in explicit chunking and counting are evidently more bias-prone than estimating via the ANS. And for addressing our initial questions, display times of 200 milliseconds are appropriate. This interval is long enough to engage the ANS (see note 8), and long enough for the use of OneToOnePlus operations; see the control experiments, noted below, in Pietroski et al. (2009). Yet it is brief enough to preclude counting and various other strategies; see note 8. One can speculate that speakers' default verification strategy takes longer to execute, because it requires other operations. But our initial questions concern (1b), (1c), and (1d).

To assess whether speakers use the ANS to approximate cardinalities in evaluating (1), given a display, the stimuli should vary with regard to the ratio of yellow to nonyellow dots—controlling for area and other factors—since performance driven by the ANS is ratio dependent, in accord with Weber's Law. Given independent results from other labs (see Dehaene 2007), there are testable predictions of how performance driven by the ANS should degrade as the ratio of yellow to nonyellow dots approaches unity.

Our method was to present native speakers of English with displays like those below for 200 ms.

[ FIGURE 1 HERE ]

23

These particular scenes include ten yellow dots and eight blue dots, making the target sentence true. In the first scene, the dots are scattered randomly. The others were generated so that each blue dot would seem to be paired with a yellow dot, leaving two yellow dots not paired with any blue dot. In the second scene, eight yellow-blue pairs are scattered, leaving two yellow dots "alone." In the third, each row contains dots of each color. In the fourth, the color in each column is uniform, with the extra yellow dots also suggesting a longer yellow line. Participants in the experiment were presented with many such displays, with varying ratios, counterbalancing for (i) whether there were more yellow dots or more blue dots, and (ii) whether or not the greater number of dots corresponded to the greater number of pixels.

Performance was unaffected by these variations.[19] For the first three trial types—Scattered Random, Scattered Pairs, and Mixed Columns—performance was as predicted by the ANS model. Indeed, the fit to the psychophysical model was quite tight (with $R^2$ values greater than .85), as shown in figure 2 and table 1. Participants were clearly using the ANS to answer the question. There was no effect of the contrast between Scattered Random trials and the other two trial types. Nor was their any evidence that participants used a OneToOnePlus algorithm, even in an error prone way; see Pietroski et al. (2009) and Lidz et al. (in press) for extended discussion of the method and data analysis.

FIGURE 2 and TABLE 1 HERE]

One might worry that any participants disposed to use a OneToOnePlus algorithm would be unable to do so in 200 ms, and that use of the ANS was simply a strategy they resorted to given the time constraints. But separate studies, reported in Pietroski et al. (2009), showed that

---

[19] On 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.

200 ms is enough time to detect and report the color of "loner" dots like the two yellow outliers in the Scattered Pairs scene above. Indeed, the new participants were *better* at this task— showing improved ratio discriminability—compared with the original group's performance on the same Scattered Pairs trials, when the task was to evaluate (1). So at least for these trials, using a OneToOnePlus algorithm would have led to good performance. And to repeat, performance in the initial experiment fit the ANS model tightly, with no symptoms of *deterioration* for the Scattered Random (non-pair-off) scenes. Participants apparently used the same cardinality-estimation strategy, across the first three trial types, even though they could have used a OneToOnePlus strategy and done quite well on at least many trials.[20]

The fourth trial type, "Sorted Columns," was interesting because participants clearly used a different and very effective strategy. Performance was nearly at ceiling, without effect of ratio. Further work is needed to know what participants were doing in these trials; perhaps they quickly became sensitive to a correlation between "perceived line length" and the truth of (1). But the important point here is that participants were *not* locked into a single strategy across all trial types. An alternative strategy can become salient and preferred to a default strategy *in certain contexts*. As stressed above, our view is not that the logical form of (1) forces speakers to base their responses on estimations and comparison of cardinalities, even in situations where some other perceptually salient aspect of the situation provides an easier way of answering the question. We offer no proposal about *why* Sorted Columns triggered use of an alternative

---

[20] One can speculate that participants were unable to use a OneToOnePlus strategy on Scattered Random scenes, and that this led them to adopt an ANS-based strategy even for Scattered Pairs and Mixed Columns scenes. But this begs the question: if 'most' is understood in correspondence terms, why *didn't* participants adopt a correspondence strategy, with a result of *increased* accuracy for pair-off cases and *decreased* accuracy for other cases? We see no reason for thinking that our stimuli would have led participants who understand 'most' in terms of correspondence to abandon a correspondence strategy.

strategy, while Scattered Pairs did not trigger use of an available alternative strategy. But the fact that participants could and did shift strategies, at least for some cases, helps bolster the point that their default strategy was *not* OneToOnePlus. It seems that a special feature of Sorted Columns scenes drove participants off using an ANS strategy, not off using a OneToOnePlus strategy.

This makes it especially striking, in our view, that performance on Mixed Columns and Scattered Pairs fit the ANS model so well. For whatever 'most' means, participants were clearly able to use their ANS system to evaluate (1) across a range of scenes. This is intriguing, though not inexplicable, if 'most' is understood in terms of cardinality comparison. For as noted above, independent evidence suggests that the ANS somehow interfaces with representations of cardinalities. But if (1) is understood in terms of one-to-one correspondence, it is hard to see how information provided by the ANS could be of use in evaluating (1). For the ANS does not generate a representation of unity, or any representation of a "minimal" difference between distinct cardinalities; cp. Leslie, Gallistel, and Gelman (2008). One can hypothesize that adult counters tacitly know (11),

$$(11) \quad \#[\Phi(x)] > \#[\Psi(x)] \text{ iff } \text{ONETOONEPLUS}[\Phi(x), \Psi(x)]$$

and that participants use their ANS to evaluate the left side of (11). But if participants represent the question posed by (1) in terms of correspondence and remainders, why do they resort to the ANS—as opposed to perception of correspondence and remainders—in evaluating (1)?

*2.3 'Most' is not 'smost'*

In considering the possibility that 'most'/MOST$[\Phi(x), \Psi(x)]$ is semantically/conceptually atomic (see section 1.2), we said that *if* this so, one might expect (1)

(1)  Most of the dots are yellow

to be understood as implying that the number of yellow dots is *significantly* greater than the number of nonyellow dots. And in any case, one might find this claim about (1) plausible.

On this view, the meaning of 'most' is *more* demanding than we have been suggesting. And this predicts that willingness to endorse (1) will decrease in a way that differs from our model, as the ratio of yellow to nonyellow dots decreases towards unity. If participants understand (1) in the more demanding fashion, their responses should indicate indifference (50% affirmative) at some ratio significantly *above* unity. Or put another way, across cases where the ANS is used to estimate and compare cardinalities, the observed performance curve should degrade towards some point on the *x*-axis significantly to the right of the origin. In cases where another system with better discriminablity is used, there should be even clearer evidence of degradation towards indifference at some ratio greater than unity. But in fact, two aspects of our data revealed that participants did *not* behave as if they understood (1) as a demanding claim according to which one extra yellow dot might not suffice to make (1) true.

First, in Sorted Columns trials, 10 yellows vs. 9 blues was judged to be an instance of most dots being yellow, in 94% of such cases. Second, on the other trial types, performance accorded with a psychophysical model for the assumption that a single extra yellow dot suffices for the truth of (1). Graphically, this can be seen in the fitted curves of Figure 2: these curves do not cross the *x*-axis, indicating chance performance on the 'most'-not-'smost' question, until the ratio approximates unity. This model predicts that participants will *tend* to answer the test question affirmatively—though the tendency may be slight—for *any* positively signed difference between the yellow and nonyellow dots. 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 (1) counts as true.

Relatedly, as noted above, if participants understood 'most' as implying *significantly more* (as opposed to *at least one more*), their accuracy should have systematically deviated from our model as the ratio of yellow to nonyellow dots decreased. The model should have become less accurate if participants understood 'most' in the more demanding way. To test for such 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). The signed deviations of these means from the psychophysics model are plotted below.

[FIGURE 3 HERE]

Differences between percent correct and the psychophysics model were centered on zero 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.[21]

*2.4 Subtraction Trumps Negation*

The logic of the main experiment discussed above was simple. In a controlled setting where a certain procedure P could be used to answer a question posed with a sentence Σ, and P reflects an initially attractive hypothesis about the logical form of Σ, let the stimuli vary in a way that would make P easier or harder to employ. If success does not vary accordingly, but rather varies in a way predicted if participants used an alternative procedure P′, this tells against the hypothesis that the logical form of Σ makes P rather than P′ the default verification procedure for Σ. And if

---

[21] Several colleagues and referees have suggested that while 'most' has the more demanding (significantly greater) meaning, some situations—e.g., votes, predictions, obligations, and displays of colored dots—lead speakers to assume that a difference of one is significant. It is hard to assess this suggestion, absent an independent characterization of when a minimal difference is significant. But if 'most' has the more demanding meaning, one wonders why the increased demand is so easily nullified, as if the *appearance* of increased demand was due to pragmatic factors.

P′ is an independently plausible candidate for the default verification procedure—i.e., if P′ also reflects a plausible hypothesis about the logical form of Σ, and the information that P′ calls for can be provided by known cognitive systems—this is some evidence that Σ has the logical form corresponding to P′, at least as opposed to P.

Lidz et al. (in press) encapsulate this line of thought as the "Interface Transparency Thesis" (ITT):

> the verification procedures employed in understanding a declarative sentence are biased towards algorithms that directly compute the relations and operations expressed by the semantic representation of that sentence.

If one focuses on the language faculty, thinking about other systems interfacing with it, ITT is another way of expressing a relatively modest point we stressed at the end of section one: the *structure* of an assembled thought makes some ways of evaluating the thought—gather certain information, by whatever means, and perform certain operations if you can—more natural than others. If one focuses on other systems, thinking about the language faculty interfacing with them, ITT reflects another (old) idea: while logical forms surely reflect linguistic constraints, they also reflect the kinds of representations and computations that humans can employ independent of special linguistic capacities.

Expression meanings allow for the construction of thoughts that can be evaluated via extralinguistic cognitive systems that are subject to their own constraints. Each such thought, **T**, must exhibit a format that is recognizable by (i) other systems, like the ANS, that can themselves provide the information required to evaluate the thought, *or* (ii) mediating systems—cognitive middlemen—that can construct a thought **T\*** that is truth conditionally equivalent to **T**, at least in the context at hand, and evaluate **T\*** via (i). Some appeal to middlemen is unavoidable given that

29

confirmation is holistic. For example, given scenes with Sorted Columns, speakers can evaluate 'Most of the dots are yellow' without estimating and comparing cardinalities.

One would like to know what speakers notice about such scenes, how they notice it, and how they exploit such scene-specific knowledge. But such questions are reminiscent of the frame problem. More generally, appeal to middlemen encodes the mystery of how complex concepts are used in complex situations. So while such appeal is unavoidable, in accounts of the many ways that sentences can be evaluated for truth/falsity, such appeal is not to be welcomed in accounts of linguistic meaning that purport to describe theoretically tractable aspects of how the human language faculty interacts with other (relatively simple) computational systems. One can adopt the hypothesis—contrary to ITT—that every case of evaluation is a holistic enterprise, and that there are no default verification strategies. But this strong claim does not follow from the more mundane point that everything one knows is potentially relevant to (dis)confirmation.

Furthermore, any "translation" from one truth-specification to another requires an initial specification: no transformations without inputs. And to learn that one truth-specification is equivalent to another, at least in certain contexts, one needs to compare outputs of some ancillary system—e.g., a line-length detector—with evaluations based on the initial specification. To learn that line length is a *good* proxy for cardinality, in certain contexts, one needs some independent assessments of how many things are in the lines. So while middlemen (and learning which procedures are equivalent in which contexts) will be interesting parts of the final story concerning how language and extralinguistic cognition interact, these interesting processes require more basic semantic specifications that provide initial/default verification procedures.

As a methodological point, ITT says that middlemen are not options of first resort, and that hypotheses about logical forms should reflect this. As a thesis about cognitive architecture,

30

ITT says that middlemen are valuable additions to a mind that could operate (less intelligently) without them. Recapitulating earlier points, our suggestion is not that meanings are verification procedures, but rather that meanings are instructions to build thoughts whose logical forms can be used as default verification procedures—so long as one abandons the idea that such procedures need to be specified in terms of experience, as opposed to abstract notions like numbers and subtraction. With this in mind, let's return to the choice between (12) and (13),

(12)  $\#[\Phi(x) \,\&\, \Psi(x)] > \#[\Phi(x) \,\&\, {\sim}\Psi(x)]$

(13)  $\#[\Phi(x) \,\&\, \Psi(x)] > \#[\Phi(x)] - \#[\Phi(x) \,\&\, \Psi(x)]$

as the logical forms for (14), our target sentence for the next experiment.

(14)  Most of the dots are blue

Ideally, one would like to let scenes vary in terms of how easy it is to "select" the nonblue dots, while also letting scenes vary (orthogonally) in terms of the ratio of blue to nonblue dots. And independent studies have revealed that given a display of suitably differing objects—e.g., dots of different colors, or lines that exhibit different orientations—the objects that share a given visual feature (e.g., being blue) can be selected and attended to, making it possible to estimate the number of such objects (Halberda, Sires & Feigenson, 2006; Nagy & Thomas, 2003; Wolfe, 1998). Unsurprisingly, the ANS cannot generate an estimate of the nonblue dots as such. Nor can disjunctive categories, like blue or red, serve as a basis for selection; see Treisman and Gormican (1988), Treisman & Souther (1985). But the ANS will generate, in parallel, estimates of the *total* number of displayed objects and two or sometimes three *subsets*, so long as each subset corresponds to a distinctive perceptual feature; see Halberda et al. (2006).

Given a simple two-color display, like figure 4a, one could well notice that the nonblue dots *are* the red ones. In a three-color display, one might notice that the nonblues are the reds

31

plus the purples. So one can imagine a middleman using information provided by the ANS to determine the number of nonblues, perhaps via *addition* in a three-color trial.

[FIGURE 4 HERE]

But any such hypothesis predicts that performance should deteriorate, holding ratio fixed, especially as the number of colors rises beyond one's capacity to independently sample each.

With five colors, any strategy that relies on estimating numerosities for each color should fail, at least if displays are presented only just long enough for selection and ANS estimation. Independent results demonstrate that human adults can only represent (on average) three to four color sets from a brief display (Halberda et al., 2006). So we presented displays like those in figures 4a-4d for only 150 ms. If the number of nonblue dots is computed by subtracting the number of blue dots from the total number of dots, performance should be unaffected by the number of colors other than blue: performance on two-color and five-color cases should be about the same. Moreover, work from other labs provides independently confirmed models of the effect of subtraction, making it possible to predict *how* accuracy should decline (as the ratio of blues to nonblues approaches unity) *if* the number of nonblue dots is computed by subtraction based on two other cardinality estimates; see Dehaene (2007).

The studies reported in Lidz et al. (in press) yielded data that fit this latter model, suggesting that (13) reflects the default evaluation procedure used. As shown in Figure 5 and Table 2, performance fit the ANS-with-subtraction model quite well (with $R^2$ around .95 or higher). There was no discernible effect of the number of colors. Across trial types, participants performed as if they used the ANS to (i) estimate the number of yellow dots directly, and (ii) estimate the number of other dots by estimating the total number of dots and subtracting the estimate for the yellow dots. Failure to use an alternative strategy in four/five-color cases is

32

perhaps not surprising. But any success at all in these cases, at 150 ms, calls for comment. If (13) does not reflect a default strategy for evaluating (14), one needs to explain how and why participants hit on this effective strategy. More importantly, though, the data suggest that participants *didn't* resort to an alternative strategy in two/three-color cases. (And recall, from experiment one, that participants are able to use an alternative strategy for certain trial types.)

[FIGURE 5 and TABLE 2 HERE]

If the task was to evaluate a thought of form (12)—or if MOST[$\Phi(x)$, $\Psi(x)$] was atomic—

(12)  $\#[\Phi(x) \& \Psi(x)] > \#[\Phi(x) \& \sim\Psi(x)]$

one might expect participants to adopt a simple strategy for two-color trials: the ANS delivers an estimate for each color; so just compare the two estimates. This would yield *better* results than estimating the total number of dots and subtracting. For even if subtraction adds no noise, the total number of dots will be larger—triggering a noisier estimate—than those of any one color. But performance on two-color trials fit the ANS-with-subtraction model, as predicted by ITT given the logical form (13).

(13)  $\#[\Phi(x) \& \Psi(x)] > \#[\Phi(x)] - \#[\Phi(x) \& \Psi(x)]$

And while (13) might seem more complex than (12), if one focuses solely on the language faculty, (13) is arguably the more natural form of thought if one focuses on the perceptual systems that regularly provide the information used to evaluate sentences.

In this context, it is also worth comparing (13) with Hackl's quantificational characterization of 'most' as 'many-est'. On his view, ignoring details not relevant here, (14)

(14)  Most of the dots are blue

has the following meaning: the dots are such that the number of blue ones exceeds the number of any other nonoverlapping subset. This condition is met if and only if the focus set (i.e., the set of

33

blues) outnumbers its complement (i.e., the set of nonblues), given the restricted domain in question (i.e., the dots). But on Hackl's view, this is because 'most' invokes quantification over *all* nonoverlapping alternatives to the focus set, its complement set included. By contrast, (13) makes do with reference to the restricted domain and focus set, using the latter's cardinality twice. Correlatively, $\#[\text{DOT}(X) \& \text{BLUE}(X)] > \#[\text{DOT}(X)] - \#[\text{DOT}(X) \& \text{BLUE}(X)]$ can be evaluated in the same way, regardless of how many colors appear in the display.

If (14) was a quantificational claim regarding every subset of the dots that does not overlap with the blues, then other things equal, one would expect the difficulty of evaluating (14) to rise as the number of colors rises. But this consideration is not decisive, even if one embraces interface transparency and our data, since it is unclear which default verification procedure is implied by Hackl's proposed semantics. It seems unlikely that any natural mind would or could compare the blue dots with *every* (or *every salient*) nonoverlapping subset of the dots. If one can judge whether the blues outnumber the reds, why bother comparing the blues to the large reds (the reds on the left, the reds near a green, etc.)? And if one had to compare the blues with each nonoverlapping set of dots, then as the number of dots rose (holding the ratio of blues to nonblues constant) the number of required comparisons would explode with just a few colors.

Likewise, if one could judge whether the blues outnumber the nonblues, why bother comparing the blues with the reds? But as noted above, the nonblues might not be selectable as such. So one might think that Hackl's semantics suggests the following default verification procedure: compare the number of blue dots to the *sum of* the cardinalities of each salient subset of the dots that does not overlap with the blues, *inferring* that this sum is at least as large as any addend. So long as the salient subclasses exhaust the nonblue dots, this procedure will work. But again, the difficulty of execution (and error rate) would presumably rise as the number of salient

34

subclasses rises, especially if those subclasses overlap. So given our data, one might hope to connect Hackl's proposed semantics with the verification procedure we suggest, treating the latter as a kind of cognitive middleman: compare the number of blue dots to the *difference between* that number and the total number of dots, *inferring* that this difference is at least as large as the cardinality of any subclass of the dots that does not overlap with the blues.

We prefer to retain interface transparency, instead of adopting a semantics that always requires this kind of inference for verification. But Hackl motivates his semantics in part by arguing that the quantificational analysis—bound up with the analysis of 'most' as 'many-est'— is attractive for reasons orthogonal to those discussed here; and we cannot assess these other arguments here. Hackl also offers evidence that in some situations, speakers asked to evaluate sentences with 'most' do perform some of the comparisons that his semantics invites. Though in our view, such evidence is not probative if the questions concern which verification procedures are the defaults. We take it as given that speakers use various strategies in various situations. For us, the question is whether available procedures are *neglected*—in circumstances where they could be used to good effect—in favor of a strategy that reflects a candidate logical form for the sentence being evaluated. If so, that is evidence in favor of the candidate in question.

**3 Extensions**

Until now, we have restricted attention to sentences in which 'most' appears as a quantificational determiner that combines with a plural internal argument—like 'dots', 'dots in France', or 'of the dots'—to form a determiner phrase that combines with a plural argument (like 'are yellow').[22] But we want to end with some remarks about (2-4),

---

[22] And note that 'most of the dots' (unlike, say, 'every dot') can sensibly combine with intuitively nondistributive predicates like 'surrounded the square', 'rained down on the square', and 'clustered inside the square'; cp. Boolos (1998), Schein (1993).

(2) I have the most dots

(3) Most of the paint is yellow

(4) Most of the dot is yellow

in the context of general questions about how lexical items can be so flexible, while still respecting substantive compositionality constraints.

*3.1 Quantificational Flexibility*

If (1) has the logical form shown in (1d),

(1) Most of the dots are yellow

(1d) $\#[\text{DOT}(X) \& \text{YELLOW}(X)] > \#[\text{DOT}(X)] - \#[\text{DOT}(X) \& \text{YELLOW}(X)]$

it is tempting to conclude that 'most' itself has the following logical form: $\#[\Phi(X) \& \Psi(X)] > \#[\Phi(X)] - \#[\Phi(X) \& \Psi(X)]$. But (2) does not mean that the number of dots the speaker has exceeds the result of subtracting that number from the number of dots. Sentence (2), in which 'most' appears as an adjective, can be true if the speaker has a plurality (but not a majority) of the dots—i.e., more dots than any other player. And in Hindi, along with various Romance languages, this meaning is expressed with a word *like* 'most' that does not also appear as a quantifier with a majority implication.

In this connection, an old and potentially relevant thought—noted above, in connection with Hackl's proposal—is that 'most' is a superlative version of 'more', which can figure in constructions like (15) and (16); cp. Bresnan (1973) and references there.

(15) More than three dots are yellow

(16) More of the dots are yellow than blue

Such examples invite appeal to a ternary concept MORE-THAN$[N, \Phi(X), \Psi(X)]$ that is analytically equivalent to $\#[\Phi(X) \& \Psi(X)] > N$. So one might specify the meanings of (15-16) as follows (cp.

Bhatt and Panchva, 2004): #[DOT(X) & YELLOW(X)] > 3; #[DOT(X) & YELLOW(X)] > #[DOT(X) & BLUE(X)]. And with regard to (2), one might suggest the following kind of meaning, letting 'Y' range over alternatives to the speaker: #[DOT(X) & HAS(SPEAKER, X)] > #[DOT(X) & HAS(Y, X)].

This would preserve important aspects of (Heim/Hackl-style) quantify-over-alternatives views, which are often motivated by examples like (2). Whether or not the English determiner is a superlative with a decomposition involving quantification over alternatives, one can describe the determiner meaning as a special case of the plurality meaning, with comparison limited to the restricted complement of a concept like [DOT(X) & YELLOW(X)]; where the cardinality of the restricted complement can be represented with #[DOT(X)] − #[DOT(X) & YELLOW(X)]. In ongoing work, we have been investigating the 'more'/'most' relation, using experimental methods like those described above. Our hope is that extending such studies to comparative morphemes, like 'er' in 'taller', will then show that the results illustrated here for the English determiner bear on a wider class of constructions.

Sentence (3) presents a different issue. At least prima facie, when speakers refer to some paint, they typically refer to some stuff that does *not* have a cardinality. Given one or more dots, those dots have a cardinality: one, two, seventeen, or whatever. In Fregean terms, some cardinal concept C[Φ(X)]—ONE[Φ(X)], TWO[Φ(X)], SEVENTEEN[Φ(X)], or whatever—can combine with DOT(X) to yield a true thought. But given some paint, it seems that no number is its cardinality. So it is unclear which (if any) "objects" fall under any concept of paint. Using underline to stress this point, the variable in PAINT(X) differs somehow from the variable in [DOT(X) & YELLOW(X)]. Correlatively, (3) is *not* an instruction for how to construct #[PAINT(X) & YELLOW(X)] > #[PAINT(X)] − #[PAINT(X) & YELLOW(X)]. Even if this representation is constructable, it does not seem to be the coherent thought expressed with (3); see 3.2 below.

37

In short, entities/satisfiers are countable things. If a word is satisfied by some entities, then some number $n$ is the number of entities that satisfy the word. So if 'red' has an extension, it has a cardinality. But a decent theory of meaning must accommodate 'red mud' along with 'red stick'. One can say that 'mud' has an *extent*: it is satisfied by some stuff s iff s is mud; and since 'red mud' is satisfied by s iff s satisfies both 'red' and 'mud', 'red' is satisfied by s iff s is red. But then one needs to say what speakers know about how the stuff that satisfies 'red' relates to the countable entities (like sticks) that satisfy 'red'? Does 'red' have an extent *and* an extension?

Similar issues arise for nouns (e.g., 'lamb') that exhibit mass/count polysemy. One doesn't usually think of 'dot' as such a noun. But examples like (4) highlight a striking fact: 'most' can combine with a singular partitive phrase to form a phrase that has nothing to do with the cardinality of (the extension of) the embedded noun. Intuitively, (4) means roughly that the majority of a certain dot's surface area is yellow. In this respect, (4) is more like (3) than (1). Examples like (17) highlight the point.

(17) Most of the only large dot that was one of the dots you liked was yellow

This suggests that detailed study of 'most'—in tandem with investigations of how sentences like (3) and (4) are evaluated in scenes where a single surface exhibits two or more colors—may also shed light on the much debated topic of how "mass meanings" are related to "count meanings."

*3.2 Mass Speculation*

We cannot review the many proposals concerning mass/count polysemy, much less offer and defend a new one. But if meanings are instructions for how to build concepts, as opposed to representation-neutral satisfaction conditions, perhaps 'most dots' and 'most paint' are instructions for how to build concepts that invite different evaluation strategies. In which case, evidence of the sort discussed above can be brought to bear on debates concerning whether: the

meanings of mass nouns are fundamental, with the meanings of count nouns somehow derived; the meanings of count nouns are fundamental, with the meanings of mass nouns somehow derived; both are special cases of a more neutral kind of semantic instruction; or these meanings are analytically independent.[23]

One can ask, for example, if humans can evaluate sentences by using an ANS-like system that can estimate "quantities" of paint (mud, etc.) in terms of area, volume, or other suitable dimensions. If so, then given independent measures of discriminability for such a system, one could go on to ask how it is related to the ANS; see Odic et al. (submitted). In particular, *if* speakers evaluate a sentence like (3) by estimating the area/volume/whatever of the yellow paint and subtracting that estimate from a corresponding estimate for the total paint in question, what is the format of such estimations? Do speakers use a default unit size, given either innately or in the context, and use the ANS to estimate the relevant *number* of units? Or will experimental results reveal, as we suspect, that speakers use a system that does not depend on cardinal (or even rational) number representations in this way?

Whatever the results, the general point is that such questions are ripe for investigation by employing known techniques from the study of perceptual psychology. And thinking in these ways can in turn suggest hypotheses about meaning that might otherwise be overlooked. Since '#' stands for a mental symbol that combines with a concept of countable things—e.g., DOT(X)—

---

[23] See Rothstein (2010) for a useful survey of options—including Link (1983), Gillon (1992), Krifka, (1989), Barner and Snedeker (2005), and Chierchia (1998a, 1998b)—and an interesting proposal of her own; see also Gordon (1985), Soja et al. (1991) and references there. Relatedly, if one thinks of meanings as instructions to build concepts, one can sharply distinguish hypotheses about the forms of the concepts assembled from any hypotheses about things to be found in the mind/language-independent world. Indeed, one might reject questions about whether 'rouge' has both an extension and an extent, in favor of more overtly psychological questions about the kinds of concepts that 'rouge' can access. We suspect that progress on the latter questions has been hampered by the idea that meanings are representation-neutral, as described in section one.

to form a concept of the *number* of things that fall under that concept, let '$Q$' stand for a mental symbol that combines with a "mass concept" like PAINT(X) to form a concept of the *quantity* that falls under that concept (cp. Cartwright (1975a,b), Koslicki (1999)); where quantities, like numbers, can be identified with equivalence classes of concepts (see notes 17 and 18). For any mass-concept, think of its quantity as the class of concepts that are equivalent to it in some way specified by that concept—mass, volume, area, or perhaps another way of measuring the extent (as opposed to the extension) of a concept; cp. Higginbotham (1994).

On this view, the meaning of (3) directs construction of the thought shown in (3a),

(3)    Most of the paint is yellow

(3a)  $Q[\underline{\text{PAINT}}(\text{X}) \,\&\, \underline{\text{YELLOW}}(\text{X})] > Q[\underline{\text{PAINT}}(\text{X})] - Q[\underline{\text{PAINT}}(\text{X}) \,\&\, \underline{\text{YELLOW}}(\text{X})]$

with 'yellow' also used to fetch a mass-concept, making 'yellow' as polysemous as 'dot'. Relatedly, one can say that (1) is not synonymous with (4), which corresponds to (4a).

(1)  Most of the dots are yellow

(4)  Most of the dot is yellow

(4a)  $Q[\underline{\text{DOT}}(\text{X}) \,\&\, \underline{\text{YELLOW}}(\text{X})] > Q[\underline{\text{DOT}}(\text{X})] - Q[\underline{\text{DOT}}(\text{X}) \,\&\, \underline{\text{YELLOW}}(\text{X})]$

If the goal is to specify truth conditions, this proposal is at best boring: it simply replaces appeal to cardinalities with appeal to quantities. But from the perspective urged here, interesting and testable questions arise immediately: can humans compare and subtract the relevant quantities; can words make contact with the requisite representations; and do words exhibit this kind of polysemy, so that replacing a count noun with a mass noun corresponds to a difference in logical form that also makes for a difference in default verification procedure, as predicted by the

Interface Transparency Thesis?[24]  We suspect that the answers are affirmative, and that in any

case, the studies used to find out will raise further interesting and testable questions, whose

answers will suggest more articulated semantic theories that are increasingly *not* representation-

neutral.

---

[24] Note that *if* cardinality-comparison for sentences like (1) shows a different behavioral profile than quantity-comparison for sentences like (3) and (4), this could be a valuable tool in the study of many other constructions. Holding a given display like Figure 1 fixed across participants, those asked to evaluate 'Most of the viruses are yellow' should show the cardinality-comparison profile, while those asked to evaluate 'Most of the virus is yellow' should show the quantity-comparison profile. In ongoing work, we are trying to test this prediction.

**References**

Barner, D., Brooks, N., & Bale, A. (2010). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, 118(1):84-93.

Barner, D., Chow, K., & Yang, S. (2009). Finding one's meaning: A test of the relation between quantifiers and integers in language development. *Cognitive Psychology, 58*, 195–219.

Barner, D., Libenson, A., Cheung, P., & Takasaki, M. (2009b). Cross-linguistic relations between quantifiers and numerals in language acquisition: Evidence from Japanese. *Journal of Experimental Child Psychology, 103,* 421-440.

Barner, D., & Snedeker, J. (2005). Quantity judgments and individuation: evidence that mass nouns count. *Cognition*, *97*(1), 41-66.

Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, *4*(2), 159–219.

Bhatt, R. and R. Pancheva (2004). Late Merger of Degree Clauses. *Linguistic Inquiry* 35:1-45.

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

Bresnan, J. (1973). Syntax of the Comparatives Construction in English. *Linguistic Inquiry* 4: 275-345.

Cantlon, J., Platt, M., & Brannon, E. (2009). Beyond the number domain. *Trends in Cognitive Sciences*, *13*(2), 83-91.

Cartwright, Helen (1975a). Some Remarks about Mass Nouns and Plurality. *Synthese* 31:395-410.

Cartwright, Helen (1975b). Amounts and Measures of Amounts. *Nous* 9:143-163.

Chierchia, G. (1998a). Plurality of mass nouns and the notion of "semantic parameter". *Events and Grammar*, *70*, 53–103.

Chierchia, G. (1998b). Reference to kinds across language. *Natural Language Semantics*, *6*(4), 339–405.

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

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1986). *Knowledge of Language*. New York: Praeger.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, N. (2000). *New Horizons in the Study of Language and Mind.* Cambridge: Cambridge University Press.

Church, A. (1941). *The Calculi of Lambda Conversion*. Princeton: Princeton University 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. (1967). Truth and Meaning. *Synthese* 17: 304-23.

Davidson, D. (1974). Belief and the Basis of Meaning. *Synthese*, 27, 300-23.

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*. New York: Oxford University Press.

Dehaene, S. (2007). Symbols and quantities in parietal cortex: elements of a mathematical theory of number representation and manipulation. In Patrick Haggard, Yves Rossetti, and

Mitsuo Kawato, editors, *Sensorimotor foundations of higher cognition, volume XXII of Attention and Performance*, chapter 24, 527-574. Harvard 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. & Carey, S. (2003). *Tracking individuals via object-files: Evidence from infants' manual search*. Developmental Science, 6, 568-584.

Feigenson, L. (2007). The equality of quantity. *Trends in Cognitive Sciences*, *11*(5), 185-187.

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

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

Fodor, J. (2003). *Hume Variations*. Oxford: OUP.

Fodor, J. and Lepore, E. (2002). *The Compositionality Papers*. Oxford: Oxford University Press.

Foster, J. (1976). Meaning and Truth-Theory. In *Truth and Meaning: Essays in Semantics*, G. Evans and J. McDowell, eds. New York, NY: Oxford University Press.

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).

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

Gallistel, C. R. (1989). Animal Cognition: The Representation of Space, Time and Number. *Annual Review of Psychology*, *40*(1), 155-189.

Gallistel, C.R. & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences*, 4(2), 59-65.
Most versus More Than Half. *Natural Language Semantics*.

Gillon, B. S. (1992). Towards a common semantics for English count and mass nouns. *Linguistics and Philosophy*, *15*(6), 597–639.

Gordon, P. (1985). Evaluating the semantic categories hypothesis: the case of the mass/count distinction. *Cognition, 20,* 209–242.

Hackl, M. (2009). On the grammar and processing of proportional quantifiers: most versus more than half. *Natural Language Semantics*, *17*(1), 63–98.

Halberda, J., Hunter, T., Pietroski, P., & Lidz, J. (submitted). The Language-Number Interface: Evidence from the acquisition of 'most'.

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213), 665-668.

Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple Spatially Overlapping Sets Can Be Enumerated in Parallel. *Psychological Science*, *17*(7), 572-576.

Halberda, J., Taing, L. & Lidz, J. (2008). The age of 'most' comprehension and its potential dependence on counting ability in preschoolers. *Language Learning and Development*, 4(2), 99-121

Heck, R. (1993). The Development of Arithmetic in Frege's *Grundgesetze der Arithmetik*. *Journal of Symbolic Logic*, 58, 579-601. Reprinted with postscript in Demopolous 1994.

Heck, R. (2000). Counting, cardinality, and equinumerosity. *Notre Dame Journal of Formal*

*Logic* 41,187–209.

Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. Wiley-Blackwell.

Heim, I. (1999). Superlatives. MIT Lecture Notes, Semantics Archive
    http://semanticsarchive.net/Archive/TI1MTlhZ/Superlative.pdf

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

Higginbotham, J. (1994). Mass and Count Quantifiers." *Linguistics and Philosophy* 17:447-480.

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

Hunter, T., Halberda, J., Lidz, J. and Pietroski, P. (2009). Beyond Truth Conditions: the
semantics of 'most'. SALT 18, Proceedings of the Semantics and Linguistic Theory Conference.

Keenan, E. and Stavi, J. (1986). A semantic characterization of natural language determiners.
    *Linguistics and Philosophy*, 9: 253-326.

Koslicki, K. (1999). The Semantics of Mass-Predicates. *Nous* 33: 46-91.

Krifka, M. (1989). Nominal reference, temporal constitution and quantification in event
    semantics. *Semantics and contextual expression*, *75*, 115.

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

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*(2), 395-438.

Leslie, A. M., Gallistel, C. R., & Gelman, R. (2008). Where integers come from. In P.
    Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind Volume 3: Foundations and
    the future.*, Evolution and cognition (pp. 109-138). New York, NY US: Oxford
    University Press.

Lidz, J., Halberda, J., Pietroski, P., & Hunter, T. (in press). Interface Transparency Thesis and
    the Psychosemantics of most. *Natural Language Semantics*.

Link, G. (1983). The logical analysis of plurals and mass terms: A lattice-theoretical approach.
    *Meaning, use and interpretation of language*, *21*, 302–323.

Ludlow, P. (2005). LF and Natural Logic. In G. Preyer and G. Peter, eds., *Logical Form*
    (Oxford: OUP).

Marr, D. (1981). *Vision: a computational investigation into the human representation and
    processing of visual information* (13th ed.). New York: Freeman.

Montague, R. (1974). *Formal Philosophy*. New Haven: Yale University Press.

Mostowski, A. (1957). On a generalization of quantifiers," *Fund. Math.*, 44: 12-36.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality.
    *Nature*, 215, 1519-1520.

Nagy A.L., & Thomas G. (2003). Distractor heterogeneity, attention, and color in visual search.
    *Vision Research, 43 (14),*1541-1552.

Nieder, A., & Miller, E. K. (2004). A parieto-frontal network for visual numerical information in
    the monkey. *Proceedings of the National Academy of Sciences of the United States of
    America*, *101*(19), 7457.

Odic, D., Pietroski, P. Hunter, T., Lidz, J., and Halberda J. (in preparation). Interface
    Transparency and the Count/Mass Distinction.

Papafragou, A. & Schwarz, N. (2005/2006). Most wanted. *Language Acquisition: A Journal of
    Developmental Linguistics*, 13(3), Special issue: On the Acquisition of Quantification.
    207-251.

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

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., Dehaene, S., et al. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, *116*(1), 33-41.

Pietroski, P. (2005). Meaning Before Truth. In G. Preyer and G. Peters (eds.), *Contextualism in Philosophy* (Oxford: Oxford University Press).

Pietroski, P. (2010.) Concepts, Meanings, and Truth: First Nature, Second Nature, and Hard Work. *Mind and Language* 25: 247-78.

Pietroski, P., Lidz, J., Hunter, T., & Halberda, J. (2009). The Meaning of 'Most': Semantics, Numerosity and Psychology. *Mind & Language*, *24:* 554-585.

Quine, W. (1951). Two Dogmas of Empiricism. *The Philosophical Review* 60: 20-43.

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

Rothstein, S. (2010). Counting and the Mass/Count Distinction. *Journal of Semantics 27*:343-97.

Soja, N., Carey, S., & Spelke, E. S. (1991). Ontological categories guide young children's inductions of word meaning: Object terms and substance terms. *Cognition,38*, 179-211.

Schein, B. (1993). *Plurals*. Cambridge, MA: MIT Press.

Tarski, A. (1933). The Concept of Truth in Formalized Languages.
Reprinted in *Logic, Semantics, Metamathematics* 2nd edition),
tr. J.H. Woodger, ed. J. Corcoran (Indianapolis: Hackett).

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: General*, *114*(3), 285-310.

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

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

Wynn, K. (1992). Children's acquisition of the number words and the counting system, *Cognitive Psychology*, *24*(2), 220–251.

Zalta, E. (2003). Frege (logic, theorem, and foundations for arithmetic).
*The Stanford Encyclopedia of Philosophy* (Fall 2003 Edition), Edward N. Zalta (ed.),
http://plato.stanford.edu/archives/fall2003/entries/frege-logic.

Zosh, J., Halberda, J., and Feigenson, L. (2011). Memory for multiple visual ensembles in infancy. *Journal of Experimental Psychology: General, 140(2)*, 141-158.

**Figures and Citations**

Figure 1: Four examples of displays from Pietroski et al. (2009). Under black and white viewing, dark grey are blue, and light grey are yellow dots. (a) "Scattered Random" display. (b) "Scattered Pairs" display. (c) "Column Pair Mixed" display. (d) "Column Pair Sorted" display.

Figure 2: Results from the experiment of Pietroski et al (2009). Each dot represents the data from the participant. The ratio is calculated by dividing the larger set by the smaller; thus, smaller ratios are more difficult to discriminate. The data was fitted to a classic Weber model attested by the literature. Each fit was excellent (all $r^2 > 0.98$). As can be seen, three of the four conditions showed identical results, with only the "Column Sorted Pairs" displaying unique performance.

Table 1: Parameter estimates from psychophysical model used in Pietroski et al (2009).

Figure 3: Data from Pietroski et al. (2009) that demonstrates that the signed error was evenly distributed in the positive and negative directions, suggesting that participants treated "most" as true if one set had more than 50% of the total number of dots.

Figure 4: Four examples of displays from Lidz et al. (in press). Under black and white viewing, dark grey are blue. (a) Two-color display, which replicates Pietroski et al (2009). (b) A three-color display (c) A four-color display. (d) A five-color display.

Figure 5: Results from the experiment of Lidz et al (in press). Each dot represents the data from the participant. The ratio is calculated by dividing the larger set by the smaller; thus, smaller ratios are more difficult to discriminate. The data was fitted to a classic Weber model attested by the literature. Each fit was excellent (all $r^2 > 0.98$). As can be seen, all conditions were comparable, suggesting that the number of sets did not matter for verifying "most".

Table 2: Parameter estimates from psychophysical model used in Lidz et al (2011).

Figure 1
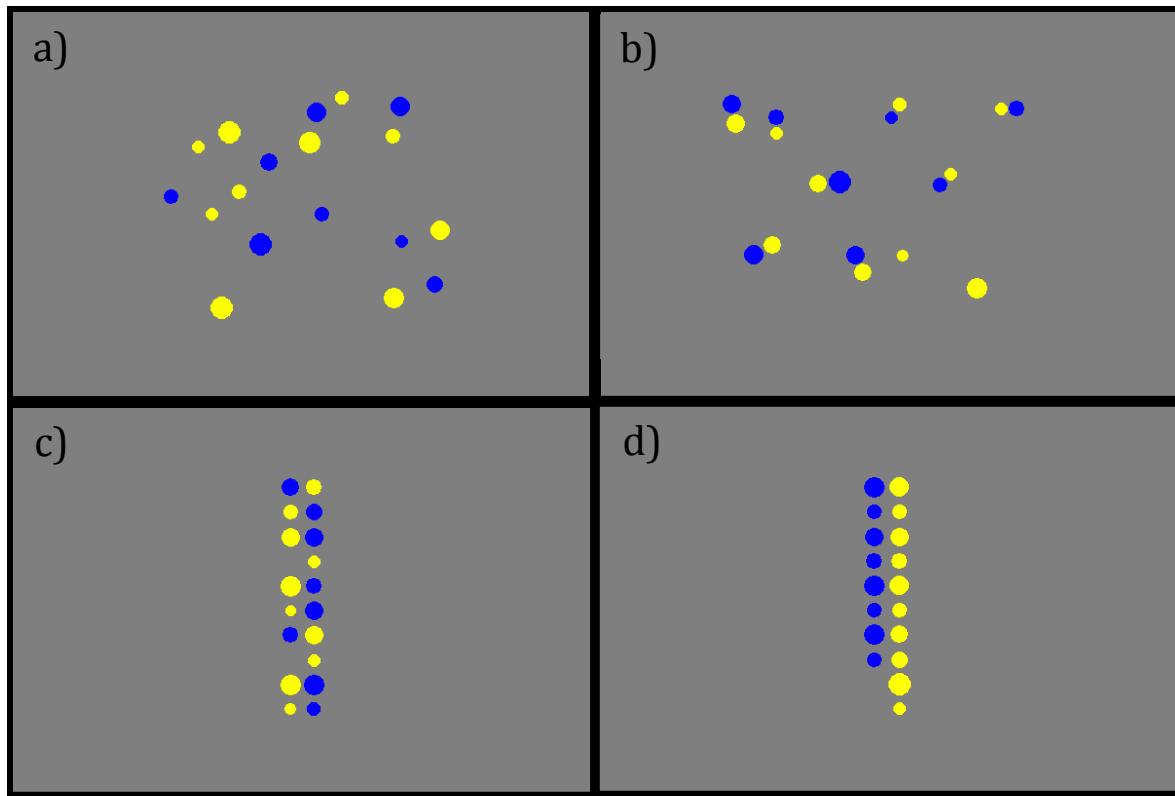
Figure 2
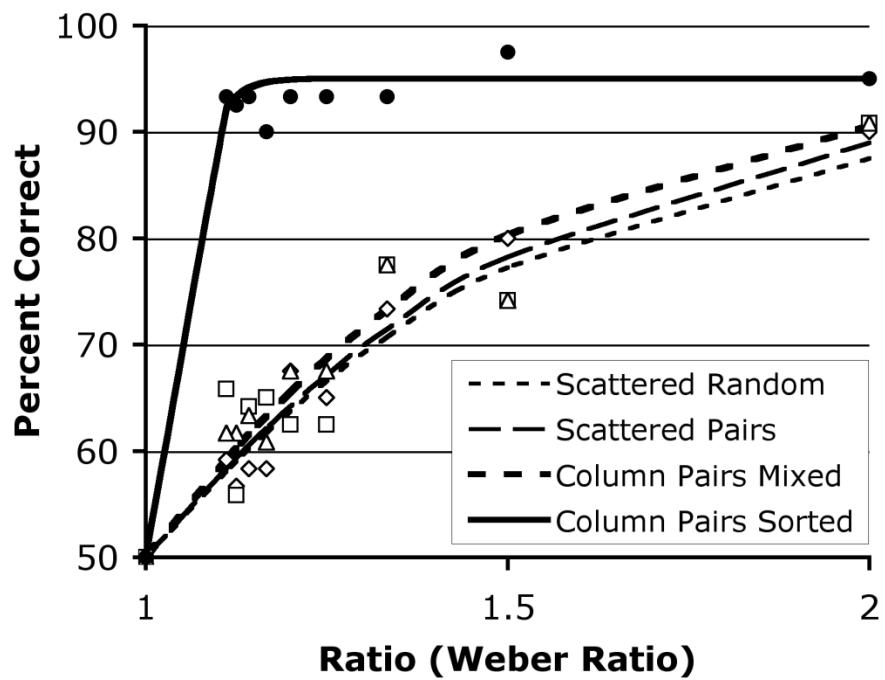
Table 1

| Trial Type | R² | Critical Weber Fraction | Nearest Whole-Number Ratio |
|---|---|---|---|
| Scatter Random | .9677 | .32 | 3:4 |
| Scatter Pairs | .8642 | .33 | 3:4 |
| Column Pairs Mixed | .9364 | .30 | 3:4 |
| Column Pairs Sorted | .9806 | .04 | 25:26 |

Figure 3
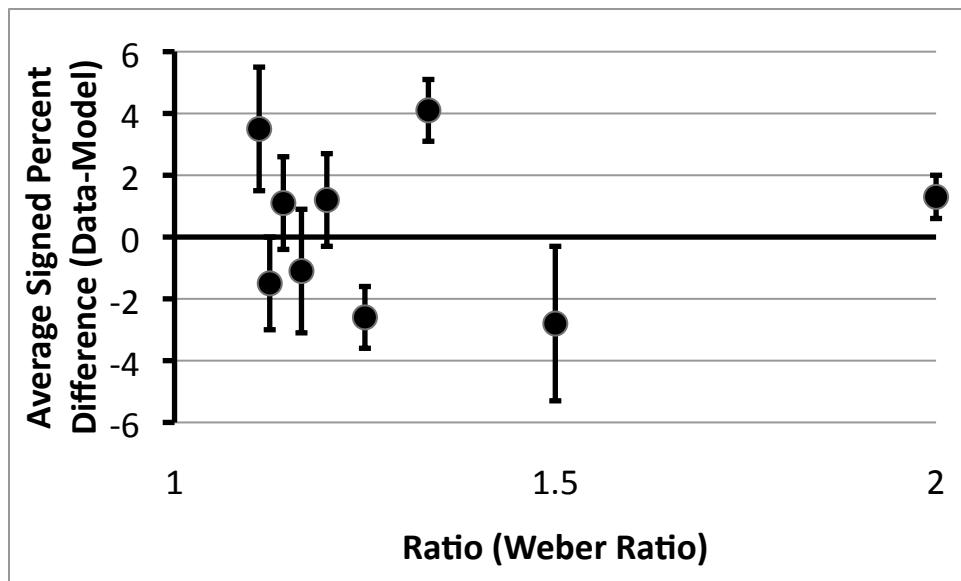
Figure 4
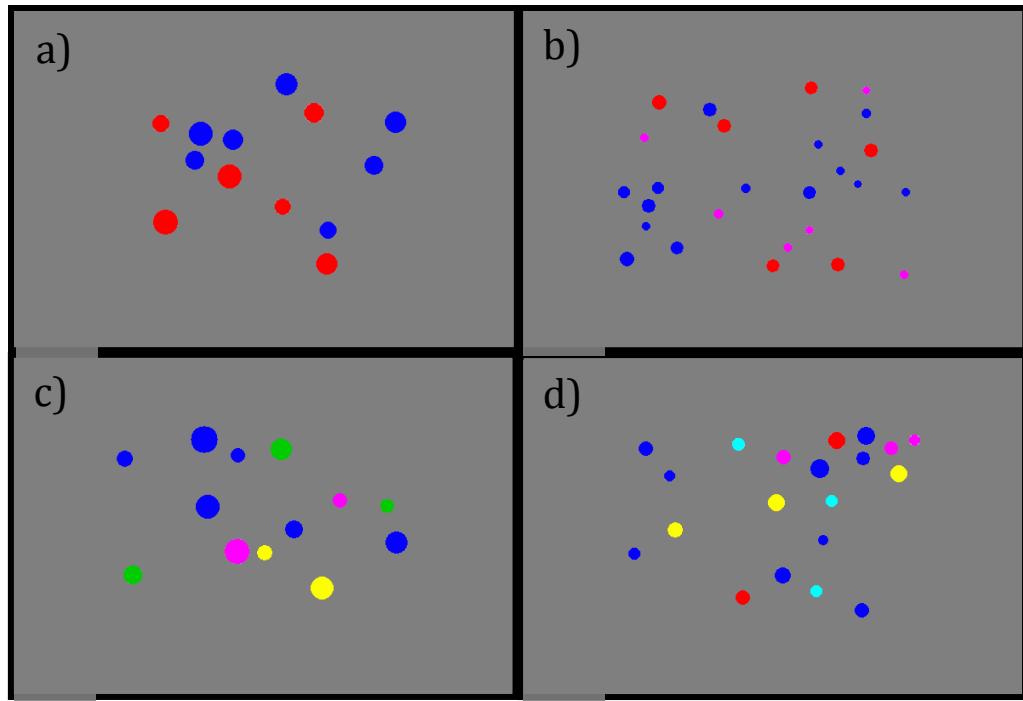
Figure 5
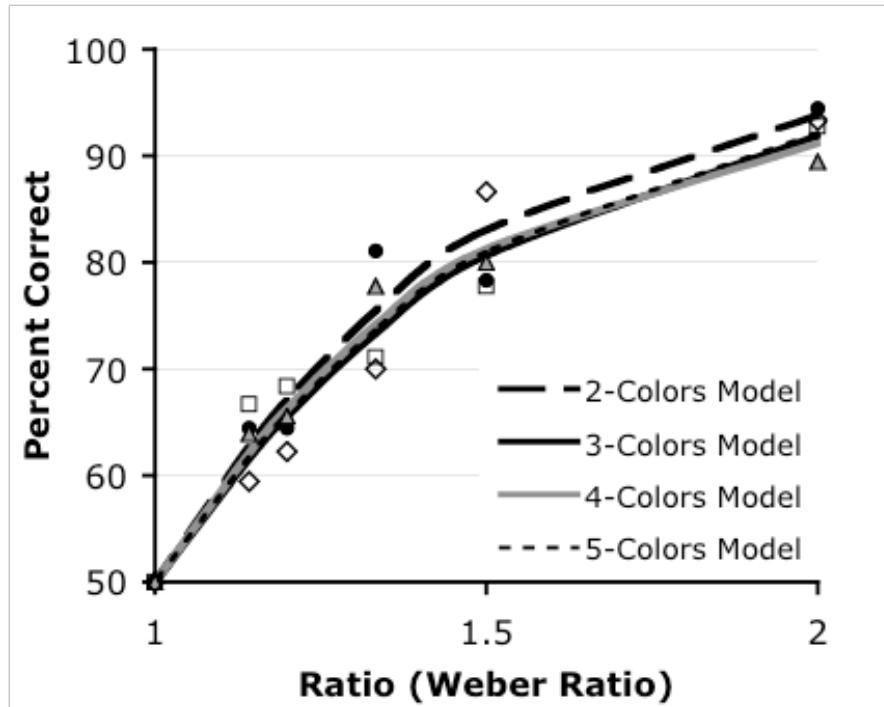
Table 2

| Trial Type | R² | Critical Weber Fraction | Nearest Whole-Number Ratio |
|---|---|---|---|
| 2-Colors | .9480 | .29 | 3:4 |
| 3-Colors | .9586 | .32 | 3:4 |
| 4-Colors | .9813 | .28 | 3:4 |
| 5-Colors | .9625 | .32 | 3:4 |