Ann. N.Y. Acad. Sci. ISSN 0077-8923

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES
Special Issue: Annals *Reports*

Original Article

Linguistic meanings as cognitive instructions

Tyler Knowlton,[1] Tim Hunter,[2] Darko Odic,[3] Alexis Wellwood,[4] Justin Halberda,[5] Paul Pietroski,[6] and Jeffrey Lidz[1]

[1]Department of Linguistics, University of Maryland, College Park, Maryland. [2]Department of Linguistics, University of California, Los Angeles, California. [3]Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada. [4]School of Philosophy, University of Southern California, Los Angeles, California. [5]Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, Maryland. [6]Department of Philosophy, Rutgers University, New Brunswick, New Jersey

Address for correspondence: Tyler Knowlton, Department of Linguistics, University of Maryland, 1413H Marie Mount Hall, 7814 Regents Drive, College Park, MD 20742. tzknowlt@umd.edu

Natural languages like English connect pronunciations with meanings. Linguistic pronunciations can be described in ways that relate them to our motor system (e.g., to the movement of our lips and tongue). But how do linguistic meanings relate to our nonlinguistic cognitive systems? As a case study, we defend an explicit proposal about the meaning of *most* by comparing it to the closely related *more*: whereas *more* expresses a comparison between two independent subsets, *most* expresses a subset–superset comparison. Six experiments with adults and children demonstrate that these subtle differences between their meanings influence how participants organize and interrogate their visual world. In otherwise identical situations, changing the word from *most* to *more* affects preferences for picture–sentence matching (experiments 1–2), scene creation (experiments 3–4), memory for visual features (experiment 5), and accuracy on speeded truth judgments (experiment 6). These effects support the idea that the meanings of *more* and *most* are mental representations that provide detailed instructions to conceptual systems.

Keywords: language; meaning; semantics; psycholinguistics; vision

Introduction

Human language connects pronunciations and meanings. For spoken languages at least, the pronunciation side of things is relatively well understood. A word like *cat* can be thought of as a collection of instructions to the motor planning system: particular details about how the lips and tongue should be positioned and how airflow should be regulated to produce the string of sounds that make up the pronunciation “cat.”[1–3]

Likewise, we can think of meanings as providing instructions to our conceptual systems. In particular, sentence meanings can be viewed as cognitive recipes: directions for how to assemble thoughts by combining concepts that can be accessed via lexical items. But it has been difficult to investigate this mentalistic conception of meaning empirically for two reasons. First, the emphasis of work in formal semantics has been on specifying conditions under which sentences are true and has abstracted away from psychological details concerning the format of the representations. Second, and more importantly, meanings need to be stated in a format that is readable by nonlinguistic cognition. Spelling out the psycho-logical details requires a rich understanding of the cognitive systems with which meanings interface. Only when we understand enough about systems, such as memory, attention, and visual cognition, can we hold those systems fixed to detect the cognitive influence of the meaning. We believe that enough is now known to make this possible for certain carefully chosen test cases. That is, we are able to ask: What sorts of instructions do linguistic meanings provide to the rest of cognition, and How detailed are these instructions?

doi: 10.1111/nyas.14618

For most expressions, the instruction offered by the meaning will severely underdetermine the contours and associations of the resulting thought. Where *cat* might bring to mind the concept of a comforting companion for one speaker, for another, it brings to mind the concept of an allergic reaction. So, while the meaning of *cat* seems to provide us enough instruction to allow speakers to talk and think about the same sort of creature, it leaves open the precise details of the thought to be built.

On the other hand, the logical vocabulary—including words like *more* and *most*—may provide a clearer window onto the nature of semantic instructions. These words lend themselves to mathematically precise description,[4] which allows for the proliferation of logically equivalent yet formally distinct specifications. Each specification constitutes an explicit psychological hypothesis about how the expressions are represented. For example, the English word *most* can be specified—and thus might be mentally represented—in various distinct ways. Consider two proposals about the meaning of the sentence given in [1]: a proportional representation given in [2], and a comparative representation given in [3].

1. *Most pianos are black*
2. #(PIANOS & BLACK) >
   #(PIANOS) - #(PIANOS & BLACK)
   ≈ *the black pianos outnumber* ***all pianos*** *minus the black pianos*
3. #(PIANOS & BLACK) >
   #(PIANOS & ∼BLACK)
   ≈ *the black pianos outnumber the* ***non-black pianos***

Both representations adequately describe the necessary and sufficient conditions of the *most*-relation. Given some pianos, both [2] and [3] will yield the same answer about the truth of the statement in [1]. So, abstracting away from *pianos* and *black*, either would be reasonable ways to describe the word *most*. And if we take both to be empirical hypotheses about what speakers know when they have acquired the meaning of the word, it is easy to see how the distinct hypotheses correspond to distinct mental states: [2] suggests that understanding a *most*-claim like [1] requires thinking about the subset of black pianos with respect to the superset of all pianos, whereas [3] suggests that it requires comparing the subset of black pianos and the subset of nonblack pianos.

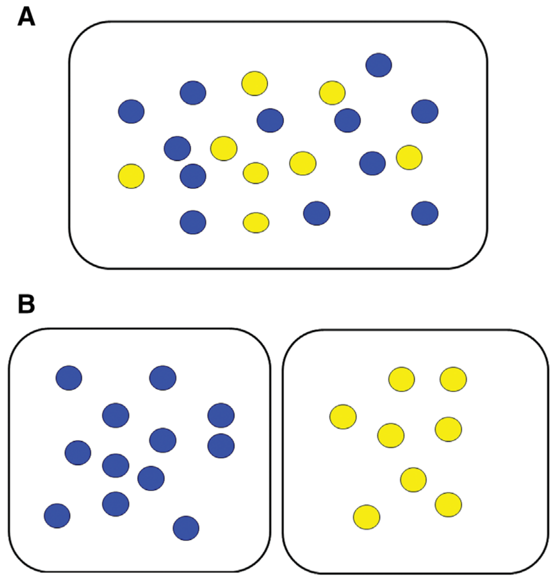

Figure 1. **Spatially intermixed (A) and spatially separated (B) displays used in experiment 1. Participants were asked to choose which image was a better example of a *more*- or *most*-sentence. They overwhelmingly preferred the spatially intermixed picture (A) when given the *most*-sentence and the spatially separated picture (B) when given the *more*-sentence. Two control groups were given the same task with *some*-sentences, either mentioning both colors or mentioning only one, but showed no such strong preference.**

The experiments reported here suggest that English-speaking adults and children understand *most* in a format that implicates a proportional relation between a subset and a superset, along the lines of [2].[4–8] This stands in contrast to their understanding of *more*, which is more along the lines of [3], at least in that *more* compares measurements of two independent subsets.[9–12] If this case study is any indication, at least some natural language meanings provide surprisingly precise constraints on thought building.

Results

Consider an image containing just blue and yellow dots, like Figure 1A or B. Given such an image, the sentences *more of the dots are blue* and *most of the dots are blue* will always pattern together with respect to truth or falsity: both statements are made true when the blue dots outnumber the yellows and false when the yellows outnumber the blues. For this reason, one might expect that any image in which blues outnumber yellows could serve as an equally good depiction of either statement. But on the other hand, while both Figures 1A and B can be visually parsed as having two subsets (blue and yellow dots), the spatially intermixed Figure 1A can be more easily parsed as a single set of dots. So, if *most* is paired with a proportional representation, as in [2], the fact that the superset is implicated might cause participants to prefer the spatially intermixed Figure 1A, in which the superset of all dots is readily apparent and easy to visually extract. And if *more* is paired with a comparative representation, similar to [3], participants may prefer the spatially separated Figure 1B, in which the two subsets are made salient and easy to visually extract. This spatial separation may hinder visual identification of the superset, but the superset is not implicated in the comparative [3].

In experiment 1, we put this question to native English-speaking adult participants. They were given two pictures—one of spatially intermixed dots (Fig. 1A) and one of spatially separated dots (Fig. 1B)—and asked to choose which was a better example of the sentence they read. Both images had exactly the same numerical distribution. However, participants ($n = 80$) preferred the spatially separated picture given the statement *there are more blue dots than yellow dots* and the spatially intermixed picture given the statement *most of the dots are blue* ($\chi^2 = 12.93$, $P < 0.001$). Here, a linguistic difference induces a visual preference, suggesting an interaction between meaning and early-to-mid level vision: the comparative (two set) or proportional (subset–superset) understanding leads to a bias for a scene that promotes a two-set (spatially separated) or subset–superset (spatially overlapping) visual parse.

To test for a potential confound of mentioning both colors in the *more*-sentences and only one in the *most*-sentence, a control group of participants ($n = 40$) was shown the same images and asked to choose which was the better example of the statement *there are some blue dots and some yellow dots* or the statement *there are some blue dots*. Contrary to the prediction that the mention of both versus one set drives preferences, we found no difference between the two groups ($\chi^2 = 1.95$, $P = 0.16$). Moreover, this nonsignificant effect numerically went in the opposite direction of the experimental condition: participants who were given the *some*-sentence that only mentioned one color were numerically more likely to choose the spatially separated image than participants who were given the *some*-sentence that mentioned both colors (26% versus 10%). This suggests that, in experiment 1, it was the mention of *more* and *most*, not the mention of one- or two-color words, that drove the result (experiments 3b, 3e, 5, and 6 directly controlled for this worry by mentioning the same number of colors in each condition).

The same preferences were expressed in the reverse inference as well: in experiment 2, participants ($n = 93$) were shown either a spatially intermixed or a spatially separated picture (Fig. S1, online only) in isolation and asked to choose which sentence better described it. Despite both sentences being truthful descriptions of either image, participants tended to pick the *more*-statement given a spatially separated picture and the *most*-statement given a spatially intermixed variant ($\chi^2 = 5.73$, $P < 0.05$). Here, visual biases induce a preference in language, despite the sentences having the same truth-conditions and containing no explicit mention of the spatial arrangement of dots.

This phenomenon is not restricted to perception. In experiment 3, participants ($n = 200$ overall; 40 per version) were given a *more*- or *most*-statement and asked to create an image that would make it true. They were allowed to place blue and yellow dots on an iPad in any way they saw fit. Participants given the statement *there are more blue dots than yellow dots* separated the two groups of dots (Fig. 2A), whereas those given the statement *most of the dots are blue* produced a single cluster of interspersed blue and yellow dots (Fig. 2B). That is, the distance between the blue and yellow clusters was significantly larger in the *more* condition than in the *most* condition ($t_{291.4} = 18.8$, $P < 0.001$). Here, language induced the same spatial bias on participants' own image creation.

This effect was found in four additional versions of the task (Figs. S2 and S3, online only). Each version made minor changes to the instructions, including, in two cases, matching the surface syntax of the statements (e.g., so that the same colors were mentioned in both conditions). Moreover, this pattern of performance is present from a young age. Experiment 4 replicated the picture creation task of experiment 3 with seventy-seven 4- to 9-year-olds (ages: 4 years, 0 months to 8 years, 10 months; mean:

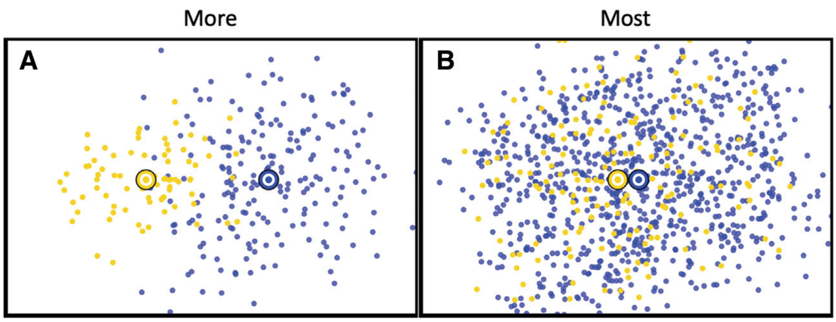

Figure 2. Results from the *more* (A) and *most* (B) conditions of one of five versions of experiment 3. Blue and yellow dots represent dots placed by participants after being asked to make it true that *there are more blue dots than yellow dots* (A) or *most of the dots are blue* (B) (see SI for other sentences tested). The dots are rotated and shifted to a common axis so that blue dots appear on the right. Large blue and yellow disks represent the centroids of these clusters ($n = 40$).

6 years, 4 months). The result is the same: children create images with two separate clusters given a *more*-statement and images with one spatially intermixed cluster given a *most*-statement ($t_{2192.4} = 19.9$, $P < 0.001$) (Figs. S2 and S3, online only).

To be sure, there is debate over when exactly learners acquire *most*.[13–15] Indeed, 17 additional children—13 of whom were between 3 and 4 years old—participated in the *most* condition but were excluded for putting 0 yellow dots on the screen. Though this is still a logically valid way to make it true that *most of the dots are blue*, it perhaps signals that these children did not understand the sentence in an adult-like way, especially given that no children in the *more* condition behaved this way. On the other hand, the fact that the 77 remaining children show the same bias as adults offers some evidence that they have an adult-like understanding of *most* and *more*. Of course, fully describing the acquisition of these words will require longitudinal work. At minimum, the results of experiment 4 show that the one- versus two-group preference found in adults emerges relatively early and does not require a lifetime of experience with *more* and *most*.

As mentioned at the outset, there is a sense in which the *more*- and *most*-sentences in these four experiments carry the same significance: that the blue dots outnumber the yellow dots. But the robust differences in preferred scene–sentence pairing suggest that these statements connect to representations that are richer than truth conditions. In particular, *most*, understood as in [2], leads participants to prefer and create images where the dots are spatially intermixed, making the superset easy to visually identify. On the other hand, *more* is understood more along the lines of [3], which does not implicate the superset but does implicate both subsets. This leads adults and children to prefer and create images where the dots are spatially separated.

In addition to scene–sentence correspondence, the proposed representations predict differences in how speakers will interrogate the world for information and what they will encode in memory. For instance, if the representation of a sentence like *more of the dots are blue* mentally highlights both the blue dots and the relevant comparison subset—the yellow dots—then, all else being equal, participants should have better memory for the yellow dots following a *more*-statement than following the minimally different *most of the dots are blue*.

Experiment 5 deployed this logic, testing 4- to 8-year-old children ($n = 250$, ages: 4 years, 0 months to 8 years, 5 months; mean: 6 years, 5 months). They were given a simple image of blue and yellow dots (Fig. 3A) and asked questions like, Did the blue team paint more/most of the dots? The comparison set was never mentioned: the only difference between conditions was the word *more* or *most*. After answering this yes/no question, the dots disappeared and participants were asked to touch on the iPad where they thought the middle of a particular set was (e.g., "Where was the middle of the yellow dots? Touch where the middle of the yellow dots was"). Because attending to and representing a group affords knowledge of its summary statistics,[16–21] we predicted that children would better remember the centroid of the sets that were highlighted in the meaning—for example, regardless of

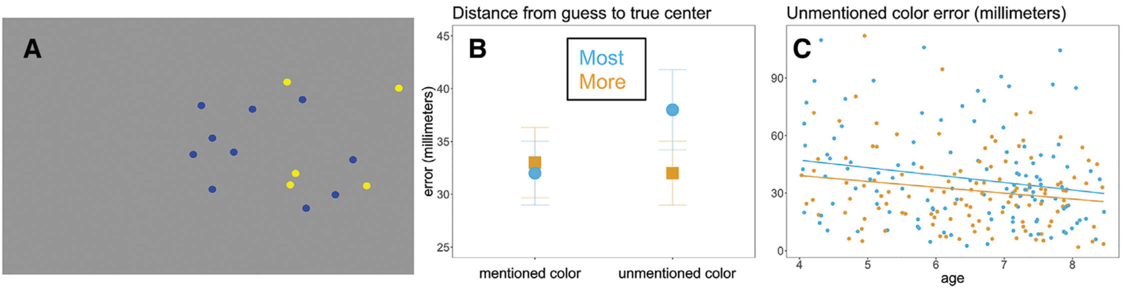

Figure 3. Example image used in experiment 5 (A), distance from actual set center when asked about the unmentioned or mentioned color following a *more*- or *most*-statement (B), and individual participants' error when asked about the unmentioned color (C). After answering whether, for example, *the blue team painted more/most of the dots*, participants saw a blank screen and were asked to touch the center of the subsets (blue and yellow). Participants in the *most* condition showed more error when asked about the unmentioned color (error bars represent 95% CI). The difference between *more* and *most* seemed to be relatively stable across the ages tested (C) ($n = 236$).

the quantifier (i.e., *more, most*), participants should remember the center of the mentioned color set (e.g., blue) reasonably well. And because the purported *more* representation in [3] highlights both subsets (even though the linguistic prompt in this case did not), participants in the *more* condition should remember the unmentioned color set (e.g., yellow) just as well. But participants in the *most* condition should have a worse memory representation of the unmentioned color set (e.g., yellow) because, according to [2], *most* requires representing the mentioned color set (e.g., blue) and the superset of all dots, but not the unmentioned color set (e.g., yellow).

Consistent with these predictions, we find that error—the distance between participants' touch and the true center of the set—is greater when asked about the unmentioned set after evaluating a *most*-statement than when asked about the unmentioned set after evaluating a *more*-statement or the mentioned set after either statement ($t_{174.03} = 2.1$, $P < 0.05$; Fig. 3B). It bears repeating that the images shown in all conditions were identical. What makes the statements true or false—the number of blue and yellow dots—is the same whether a participant is evaluating a *more*- or *most*-statement. And only one color was ever mentioned in the question/statement. For these reasons, nothing about the physical stimulus, or aspects of the sentences other than *more*/*most*, could have led participants to remember the center of mass of the unmentioned set better following *more*-statements than following *most*-statements. The difference in performance, then, must arise due to how the two statements are represented: the *more*-statement instructs participants to directly compare the blue and yellow dots, whereas the *most*-statement biases them to compare the blue dots to the superset of all dots.

Another consequence of gathering and using different information during evaluation is that participants should show distinct patterns of performance when asked to verify *more*- versus *most*-statements. For example, imagine seeing a display of 22 blue dots and 9 yellow dots, as in Figure 4A. If the display is flashed so briefly that counting is impossible, humans will rely on their Approximate Number System to approximate the cardinalities of the sets.[22,23] One hallmark of this system is that the representations it produces are "noisy," and this noise linearly increases as the numerosity being represented increases. The more numerous the group, the more uncertainty in the representation of its cardinality. In this example, the yellow dots (cardinality: 9) are represented with the most precision, the blue dots (cardinality: 22) are represented with more variability, and the superset of all dots (cardinality: 31) is represented with the most uncertainty of all three.

Now, imagine being asked whether *more of the dots are blue* or *most of the dots are blue*. Given the representations in [2] and [3], both statements are understood as comparisons between the blue dots and another set. But while the *more*-statement calls for comparing the blue dots to the yellow dots directly (i.e., $22 > 9$), the *most*-statement calls for comparing the blue dots to the result of a subtraction (i.e., $22 > 31–22$). In either case, result is the same ($22 > 9$), but in the *most*-case, more uncertainty is introduced into the right side of the

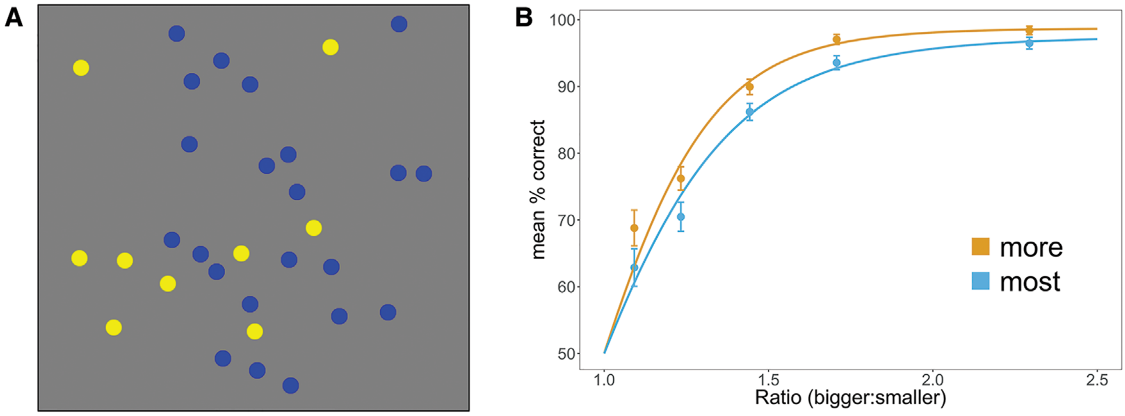

Figure 4. **Participants in experiment 6 saw partially intermixed displays of blue and yellow dots (A) presented for 200 milliseconds. The group that evaluated the *more*-statement showed better accuracy at every level of difficulty (ratio) than the group that evaluated the *most*-statement (B), despite the stimuli and the correct answer being identical in either case. Error bars represent SE, and lines represent psychophysical model fits ($n = 68$).**

inequality due to the subtraction and the larger cardinalities (31–22 versus 9). For any such display, the direct blue–yellow comparison invited by *more* will involve less uncertainty than the more indirect and noisy blue–total comparison suggested by *most*. All else being equal, then, participants should be more accurate at evaluating *more*-statements compared with *most*-statements when faced with identical images and limited viewing time.

This prediction is borne out in experiment 6. Adult participants ($n = 68$) were shown 100 partially intermixed dot displays, designed to make the direct comparison and the proportional strategies equally viable (Fig. 4A). Each display appeared for 200 milliseconds. Participants' task was simply to judge one of the following statements as true or false: *the blue team painted more/most of the dots*. As predicted, both groups of participants showed the ratio-dependent performance characteristic of the Approximate Number System (Fig. 4B; $F(4,264) = 149.5$; $P < 0.001$). And, as predicted, those in the *more* condition outperformed those in the *most* condition ($F(1,66) = 10.9$; $P < 0.01$) even though the images, the correct answers, and the surface syntax were exactly the same in either case. The only difference was the presence of the word *more* or *most* in the statement.

In experiment 6, we see not only that participants did something different when faced with *more*- or *most*-statements, but also that *most*-statements led to suboptimal performance. Participants could have gotten the right answer in either case by comparing blue dots to yellow dots directly or by comparing the blue dots to the superset of all dots. As far as the visual system is concerned, a direct-comparison strategy is clearly superior given these displays, as evidenced by performance following *more*-statements (Fig. 4B). But participants' understanding of the statement and the information they choose to gather during evaluation depends on subtle differences in the way that these words are mentally represented. Participants use the sets highlighted by these mental representations, even when the visual system offers superior alternatives. This subtle difference in understanding motivates participants to rely on different verification strategies, and, when used in this way, evidence of differences in verification strategies can be used as evidence for differences in understanding.

Discussion

The idea that subtle differences in meaning arise based on word choice is a familiar one. Certain situations may call for different ways of expressing the same truth-conditional content. A scene of a fox and a rabbit running might be described as a *chasing* or a *fleeing*, depending on whose perspective the speaker intends to highlight.[24] Likewise, a scene where some blue dots outnumber some yellow dots might be described with *more* or *most*, depending on which sets the speaker intends to highlight.

In other words, while the sentences *more of the dots are blue* and *most of the dots are blue* can be used to describe the same aspects of a speaker's environment, they nonetheless have distinct mental correlates: representation of a subset–subset relationship (*more*) or of a subset–superset relationship (*most*). These correlates have behavioral consequences, the consistency of which—across both adults and children, using a variety of different probes—is surprising because there are many logical possibilities that any individual speaker could, in principle, have considered when learning these words. For example, *most* could have been understood in a way that implicates the superset or in a way that calls for a comparison of subsets, or even one-to-one correspondence.[5,25] And the same can be said of *more*. Different speakers might, in principle, settle on different representations to pair with these words and never disagree in conversation. But despite this, we find evidence across participants and tasks that speakers have a representation for *most* like [2] and a distinct representation for *more* that is more in line with [3].

To be sure, the results presented here are between–subject comparisons, so they do not license the strong conclusion that every speaker of English understands *most* along the lines of [3]. It might instead be that merely most speakers understand *most* in this way. Future within-subject versions of these tasks will aim to settle the question empirically. But regardless of the proportion, if (at least a majority of) speakers understand a word like *most* in the same way at a fine-grained level of detail, this raises acquisition questions of the sort highlighted by Quine's Gavagai problem.[26] Many mental representations could underlie a word like *most*, presenting learners with many alternative hypotheses. And given their truth-conditional equivalence, simply noticing the truth or falsity of any number of *most*-claims will never be sufficient for a learner to rule out those alternatives. While the details of this language acquisition question are left for future work, we suspect potential answers will require that learners are equipped with substantial constraints on the space of possible representations[27–29] and an ability to learn from more subtle properties of their input than truth and falsity.[24,30]

Perhaps many, or even most, of our lexical concepts are atomic and innate.[31] The lexical item *cat* might correspond to an atomic concept CAT, which does not have the concept ANIMAL as a part. This would make CAT and ANIMAL as logically independent as CAT and TOMATO. However, given the results reported above, it seems unlikely that the members of the logical vocabulary—in particular, quantificational expressions like *most*—correspond to atomic concepts. Our results suggest that speakers represent sentences like *most of the dots are blue* in a particular format. This format implicates other logical concepts (e.g., cardinality, comparison, subtraction, superset, and subset), some of which are shared by other quantificational expressions (e.g., *more of the dots are blue*). This leaves room for debate (and future work) concerning what the logically primitive concepts are, and which, if any, vocabulary items express these concepts. Uncovering such primitives will play an important role in settling the learning question raised above. In the meantime, though, the experiments presented above support a specific proposal about the meaning of *most* that implicates notions of subset–superset comparison and subtraction.

Jerry Fodor once remarked that "it's an iron law of cognitive science that, in experimental environments, definitions always behave exactly as though they weren't there."[31] In defiance of this iron law, the six experiments presented above reveal apparent differences in the semantic structure of *more* and *most*, as represented by speakers of English. We think that Fodor's law has been overturned, in this case, for two reasons. First, we considered elements of logical vocabulary, where alternative representational hypotheses are readily available and easily testable. Second, we relied on a rich understanding of the interfacing extralinguistic systems (e.g., visual short-term memory, ensemble representations, and the Approximate Number System). This strategy makes it possible to go beyond studying the informational content of expressions (e.g., the conditions under which they are true), and to focus on precise details of the psycho-semantic representations underlying linguistic knowledge and use.

Materials and methods

Guidelines for testing human research participants were followed as certified by the Johns Hopkins University and the University of Maryland Institutional Review Boards. Participants in all experiments gave informed consent prior to viewing any study materials, and subjects' rights were protected throughout.

Experiment 1

One hundred twenty participants were recruited on Amazon Mechanical Turk. Each participant was paid \$0.20 for completing the brief survey. Participants were required to have a 90% or greater lifetime approval rating and an IP address from the United States, and they confirmed before starting the task that they were born and raised in the United States and that they speak English better than any other language.

Following an English proficiency screener, participants received three trials of a "picture selection task" where they were asked to read a sentence and select which picture best exemplified it. They were asked to make their decisions quickly and told that we were interested in their initial intuitions. After two filler trials ("Mary lit some candles on June's birthday cake" and "Two boys held three balloons"), participants were shown the critical trial (a *more-* or a *most*-sentence).

On the critical trial, one group of 40 participants saw the two images in Figure 1. The order of these images was balanced across participants. Half of those 40 participants read the sentence *most of the dots are blue* and the other half read *there are more blue dots than yellow dots*. Another group of 40 participants received the same survey with one difference: the images in Figure 1 were mirror reversed. This was to control for any potential effect of the blue being on the left in the spatially separated image. These two groups of participants were collapsed across in the analysis reported in the main text, as performance was identical in both groups. A third group of 40 participants served as the control group. They received the same survey as the first group, with one difference: on their critical trial, half read *there are some blue dots* and the other half read *there are some blue dots and some yellow dots*. As noted above, this was to control for the fact that the *most* condition mentioned only one color, while the *more* condition mentioned both blue and yellow.

Experiment 2

Ninety-three Johns Hopkins University undergraduates were recruited through the Psychological & Brain Sciences subject pool. Each student received course credit for completing a paper-and-pencil survey. For each item on the survey, they were asked to report their "gut feeling" about which of the two sentences better exemplified the picture they saw. They were instructed as follows: "*Both* of the sentences are true descriptions of the picture. Throughout this Questionnaire, we want you to choose which one of the two sentences is a better descriptor of the picture according to your own intuitions." Four filler trials were presented before the critical *more-* versus *most*-trial. These filler questions and the critical question are shown in Figure S1 (online only). The critical question had two conditions: Half of the participants saw a spatially separated picture of gray and black dots, whereas the other half saw a spatially intermixed image.

Experiments 3 and 4

Two hundred undergraduate students (recruited at Johns Hopkins University and the University of Maryland) and ninety-four 4- to 9-year-olds (recruited at Johns Hopkins University) were given an iPad running an app that provided a blank gray canvas on which blue or yellow dots could be placed. They were asked to put dots on the screen to make a *more-* or *most*-sentence true. For the adult version, whether the experimenter demonstrated how to place dots on the screen and the exact wording of the critical sentences differed in each version of the experiment (details are given in Figs. S2 and S3, online only). For the child version, all participants heard the same instructions: "Make it so that there are more blue dots than yellow dots" or "Make it so that most of the dots are blue." As noted above, 17 children on the younger end of the age range were excluded from the *most* condition for putting 0 yellow dots on the screen.

In order to normalize the dot displays produced by participants, we determined, for each participant's image, the center of the blue dots, the center of the yellow dots, and the center of the superset of all dots. Dots were rotated until the center of the blue dots was on the right and the center of the yellow dots was on the left. Then, they were repositioned so that the center of the superset corresponded with the center of the display (Fig. 2; and Figs. S2 and S3, online only).

A measure of distance between clusters—average silhouette width—was then computed for each condition (*more, most*) in each version of the task (Fig. S3, online only). For any given dot, a large positive value indicates that the dot in question is well clustered with respect to dots of the same color. A large negative value indicates that the dot in question is more likely to be a member of the other color group. A value near zero indicates that the dot in question would be an equally good fit for either color group. To the extent that the average silhouette width is higher in the *more* condition than the *most* condition, we can conclude that the dots placed by participants given a *more*-statement are more highly separated into distinct clusters compared with dots placed by participants given a *most*-statement. To this end, we conducted Welch two sample two-tailed t-tests comparing the silhouette widths in both conditions. In each version, the average silhouette width in the *more* condition was greater than that in the *most* condition (**3a**: $t_{291.4} = 18.8$, $P < 0.001$; **3b**: $t_{190.8} = 14.2$, $P < 0.001$; **3c**: $t_{537.9} = 18$, $P < 0.001$; **3d**: $t_{778.4} = 3.68$, $P < 0.001$; **3e**: $t_{635.3} = 4.41$, $P < 0.001$; **4**: $t_{2192.4} = 19.9$, $P < 0.001$).

Experiment 5

Three hundred thirty-six 4- to 8-year-olds (recruited at Johns Hopkins University and the University of Maryland) participated. Of these, 86 were excluded before any analysis for needing to tap the screen multiple times to record their guess (70), failing to understand the task (8), not being a native English speaker (2), equipment failure (4), or parental/sibling interference (2). Of the remaining 250 participants, 7 were then excluded from further analysis for answering the initial yes/no question incorrectly. Seven participants were then removed because their guess error was greater than 3 standard deviations above the mean. This left 236 participants, 118 of whom were in the *more* condition and 118 in the *most* condition.

Children were presented with a blank gray screen while the experimenter said: "I'm going to show you a picture where the blue team painted dots and the yellow team painted dots. I need you to say whether the {blue/yellow} team painted {more/most} of the dots." The experimenter tapped on the screen revealing the dot display in Figure 3A or the mirror reversed image and repeated "Did the {blue/yellow} team paint {more/most} of the dots?" until the child answered. They were encouraged regardless of their response and then shown another gray screen, at which point the experimenter asked the participant "Do you remember where the {blue/yellow} dots were? Where was the middle of the {blue/yellow} dots? Can you touch were the middle of the {blue/yellow} dots was?"

Experiment 6

Adult participants ($n = 68$) were recruited through the University of Maryland Linguistics subject pool. Each student received course credit for their participation. Their task was to judge a statement as true or false by pressing the T or F key on the keyboard. Each participant read one of four statements: *the {blue/yellow} team painted {more/most} of the dots*. Before the task began, they were given two practice trials in which the dots stayed on screen for 1 s, then two practice trials in which the dots stayed on screen for 250 milliseconds.

Displays consisted of between 8 and 25 equal-sized dots of each color. Because one color always had more dots than the other, there was a minimum of 17 dots and a maximum of 49 dots on the screen in any given trial. The easiest ratio of the "winning" color to the "losing" color was 3.125 (e.g., 25 blue dots and 8 yellow dots); the hardest ratio was 1.04 (e.g., 25 blue dots and 24 yellow dots). Dots were placed on the screen so that the displays would appear partially spatially intermixed: blue dots were restricted to the right 80% of the display window, whereas yellow dots were restricted to the left 80% of the display window (Fig. 4).

The statistical comparisons reported above are the result of a 5 (ratio bin) × 2 (*more* or *most*) ANOVA. Ratio bins were specified as follows: 1.04–1.149, 1.15–1.3125, 1.3126–1.5625, 1.5626–1.91, and 1.92–3.125. We excluded any trials in which participants took longer than 10 s to respond; however, this accounted for only 2 out of 6800 total trials. Removing these two trials, participants took an average of 1254 ms to respond in the *most* condition and an average of 1219 ms to respond in the *more* condition. This difference was not significant ($t_{328.97} = 0.922$, $P = 0.36$), confirming that the results do not reflect a speed-accuracy trade-off.

Although model fits were not used for statistical analysis, they are provided in Figure 4B. These fits were obtained using the standard model of Approximate Number System representations: Gaussian activations situated on a mental "number line" whose standard deviation increases linearly with the mean;[22,23] see the appendix of Ref. 7 for a tutorial. The rate of increase of the standard deviation is represented by the single free parameter w, so the Approximate Number System representation of some numerosity, n_1, can be modeled as a Gaussian random variable: $ANS_{n1} \sim \mathcal{N}(n_1, w \times n_1)$.

Comparison between two Approximate Number System representations, ANS_{n1} and ANS_{n2}, is then modeled as Gaussian subtraction $ANS_{n1} - ANS_{n2}$. The resulting distribution has mean $n_1 - n_2$ and standard deviation $\sqrt{(w \times n_1)^2 + (w \times n_2)^2}$. Intuitively, density of this distribution on the left of 0 indicates the likelihood of responding that n_2 is greater than n_1, whereas density to the right of 0 indicates the likelihood of responding that n_1 is greater than n_2.

The rate of correct responses, then, is $\frac{1}{2} erfc(n_1 - n_2/\sqrt{2}w\sqrt{n_1^2 + n_2^2})$, where *erfc* is the complementary error function of a Gaussian. Using this equation and maximum likelihood estimation, a single w was fit to the raw data for participants in the *more* condition ($w = 0.2$) and a second w was fit for participants in the *most* condition ($w = 0.24$).

We then added a second parameter to this model: guess percentage, g. Intuitively, g represents any reason besides the acuity of the Approximate Number System that might have led participants to answer incorrectly. On the easiest trials (a ratio of 25:8), for example, the model predicts near perfect performance. But participants might sometimes press the wrong button or blink, reducing their ceiling level of performance even on ordinarily easy trials. The parameter g is meant to capture this reduced ceiling.

In the two-parameter model, the rate of correct responses is computed the same way as in the original model, but it is multiplied by $1 - g$ (i.e., the percentage of time participants are answering using their Approximate Number System) and $\frac{g}{2}$ is added to this product (to account for the fact that only half of the error function is being plotted). Using this updated equation and maximum likelihood estimation on the binned data, w and g were fit for participants in the *more* condition ($w = 0.18$, $g = 2.5\%$) as well as for participants in the *most* condition ($w = 0.22$, $g = 5\%$). These parameters were used to generate Figure 4B.

Acknowledgments

We thank two anonymous reviewers, as well as Alexander Williams and Zoe Ovans, for their helpful comments on previous drafts, and Robert Eisinger for his help programming experiments 3 and 4. This work was supported by a grant from the James S. McDonnell Foundation on the Nature and Origins of the Human Capacity for Abstract Combinatorial Thought to J.H., P.P., and J.L., and by grants from the National Science Foundation (Doctoral Dissertation Research Improvement Grant #BCS-2017525 to J.L. and T.K., and NRT Award #DGE-1449815 to T.K.).

Author contributions

D.O., A.W., T.H., J.H., P.P. & J.L. designed experiments 1 and 2 and D.O. & J.H. collected the data. D.O., J.H. & J.L. designed experiments 3 and 4 and D.O., A.W. & J.H. collected the data. T.K., J.H. & J.L. designed and collected data for experiment 5. T.H., J.H., J.L. & P.P. designed experiment 6 and T.K. & J.L. collected the data. D.O. & T.K. analyzed the data with input from J.L. & J.H. T.K. & J.L. wrote the manuscript with input from D.O., A.W., T.H., P.P. & J.H.

Supporting information

Additional supporting information may be found in the online version of this article.

Figure S1. Filler questions (1–4) and target question (5) from experiment 2.

Figure S2. Results from experiments 3–4.

Figure S3. Silhouette width for each blue and yellow dot in each version of experiments 3–4.

Competing interests

The authors declare no competing interests.

References

1. Halle, M. 2003. *From Memory to Speech and Back: Papers on Phonetics and Phonology*, 1954–2002. The Hague: Walter de Gruyter.
2. Liberman, A.M. & I.G. Mattingly. 1985. The motor theory of speech perception revised. *Cognition* **21:** 1–26.
3. Poeppel, D.W., W. Idsardi & V. van Wassenhove. 2008. Speech perception at the interface of neurobiology and linguistics. *Philos. Trans. R. Soc. Lond. B* **363:** 1071–1086.
4. Barwise, J. & R. Cooper. 1981. Generalized quantifiers and natural language. In *Philosophy, Language, and Artificial Intelligence*. J. Kulas, J.H. Fetzer & T.L. Rankin, Eds.: 241–301. Dordrecht: Springer.
5. Pietroski, P., J. Lidz, T. Hunter & J. Halberda. 2009. The meaning of 'most': semantics, numerosity and psychology. *Mind Lang.* **24:** 554–585.

6. Hackl, M. 2009. On the grammar and processing of proportional quantifiers: most versus more than half. *Nat. Lang. Semant.* **17:** 63–98.
7. Lidz, J., P. Pietroski, J. Halberda & T. Hunter. 2011. Interface transparency and the psychosemantics of most. *Nat. Lang. Semant.* **19:** 227–256.
8. Tomaszewicz, B. 2011. Verification strategies for two majority quantifiers in Polish. *Proc. Sinn Bedeutung* **15:** 597–612.
9. Heim, I. 2000. Degree operators and scope. *Semant. Linguist. Theory* **10:** 40–64.
10. Wellwood, A. 2019. *The Meaning of More vol. 12 of Oxford Studies in Semantics and Pragmatics*. Oxford University Press.
11. Odic, D., P. Pietroski, T. Hunter, *et al.* 2013. Young children's understanding of "more" and discrimination of number and surface area. *J. Exp. Psychol.: Learn. Mem. Cogn.* **39:** 451–461.
12. Odic, D., P. Pietroski, T. Hunter, *et al.* 2018. Individuals and non-individuals in cognition and semantics: the mass/count distinction and quantity representation. *Glossa* **3:** 1–20.
13. Halberda, J., L. Taing & J. Lidz. 2008. The development of "most" comprehension and its potential dependence on counting ability in preschoolers. *Lang. Learn. Dev.* **4:** 99–121.
14. Sullivan, J., A. Bale & D. Barner. 2018. Most preschoolers don't know most. *Lang. Learn. Dev.* **14:** 320–338.
15. Papafragou, A. & N. Schwarz. 2006. Most wanted. *Lang. Acquisit.* **13:** 207–251.
16. Ariely, D. 2001. Seeing sets: representation by statistical properties. *Psychol. Sci.* **12:** 157–162.
17. Chong, S.C. & A. Treisman. 2003. Representation of statistical properties. *Vision Res.* **43:** 393–404.
18. Halberda, J., S.F. Sires & L. Feigenson. 2006. Multiple spatially overlapping sets can be enumerated in parallel. *Psychol. Sci.* **17:** 572–576.
19. Burr, D. & J. Ross. 2008. A visual sense of number. *Curr. Biol.* **18:** 425–428.
20. Alvarez, G.A. 2011. Representing multiple objects as an ensemble enhances visual cognition. *Trends Cogn. Sci.* **15:** 122–131.
21. Whitney, D. & Y. Leib. 2018. Ensemble perception. *Annu. Rev. Psychol.* **69:** 105–129.
22. Feigenson, L., S. Dehaene & E. Spelke. 2004. Core systems of number. *Trends Cogn. Sci.* **8:** 307–314.
23. Dehaene, S. 2011. *The Number Sense: How the Mind Creates Mathematics*. Oxford University Press.
24. Gleitman, L.R. 1990. The structural sources of verb meanings. *Lang. Acquisit.* **1:** 3–55.
25. Piaget, J. 1965. *The Child's Conception of Number*. Norton.
26. Quine, W.V. 1970. On the reasons for indeterminacy of translation. *J. Philos.* **67:** 178–183.
27. Chomsky, N. 1971. *Problems of Knowledge and Freedom*. Pantheon Books.
28. Gallistel, C.R. 2010. Learning organs. In *Chomsky Notebook*. J. Bricmont & J. Franck, Eds.: 193–202. Columbia University Press.
29. Lasnik, H. & J. Lidz. 2017. The argument from the poverty of the stimulus. In *The Oxford Handbook of Universal Grammar*. I. Roberts, Ed.: 221–248. Oxford University Press.
30. Lidz, J. & A. Gagliardi. 2015. How nature meets nurture: universal grammar and statistical learning. *Annu. Rev. Linguist.* **1:** 333–353.
31. Fodor, J.A. 1998. *Concepts: Where Cognitive Science Went Wrong*. Oxford University Press.