Linguistics and Philosophy (2022) 45:911–941
https://doi.org/10.1007/s10988-021-09337-8

ORIGINAL RESEARCH

The mental representation of universal quantifiers

Tyler Knowlton[1] · Paul Pietroski[2] · Justin Halberda[3] · Jeffrey Lidz[1]

Accepted: 10 June 2021 / Published online: 8 September 2021

Abstract A sentence like *every circle is blue* might be understood in terms of individuals and their properties (e.g., for each thing that is a circle, it is blue) or in terms of a relation between groups (e.g., the blue things include the circles). Relatedly, theorists can specify the contents of universally quantified sentences in first-order or second-order terms. We offer new evidence that this logical first-order vs. second-order distinction corresponds to a psychologically robust individual vs. group distinction that has behavioral repercussions. Participants were shown displays of dots and asked to evaluate sentences with *each*, *every*, or *all* combined with a predicate (e.g., *big dot*). We find that participants are better at estimating how many things the predicate applied to after evaluating sentences in which universal quantification is indicated with *every* or *all*, as opposed to *each*. We argue that *every* and *all* are understood in second-order terms that encourage group representation, while *each* is understood in first-order terms that encourage individual representation. Since the sentences that participants evaluate are truth-conditionally

✉ Tyler Knowlton
tzknowlt@umd.edu

Paul Pietroski
paul.pietroski@rutgers.edu

Justin Halberda
halberda@jhu.edu

Jeffrey Lidz
jlidz@umd.edu

[1] Department of Linguistics, University of Maryland, 7814 Regents Drive, College Park, MD 20742, USA

[2] Department of Philosophy, Rutgers University, 106 Somerset St. (5th Floor), New Brunswick, NJ 08901, USA

[3] Department of Psychological and Brain Sciences, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA

equivalent, our results also bear on questions concerning how meanings are related to truth-conditions.

Keywords Psychosemantics · Quantification · Semantics-cognition interface · Universal quantifiers

1 Introduction

Lexical items are often said to have meanings that determine extensions, at least relative to contexts.[1] In particular, it's often useful to assume that quantificational expressions like *every* have extensional "semantic values," and that various words for universal quantification—including *every*, *each*, and *all*—have a shared semantic value that can be specified equally well in many ways, including (1–4).

(1) $\lambda\Psi \cdot \lambda\Phi \cdot \forall x{:}\Psi(x)[\Phi(x)]$
(2) $\lambda\Psi \cdot \lambda\Phi \cdot \sim \exists x{:}\Psi(x)[\sim \Phi(x)]$
(3) $\lambda\Psi \cdot \lambda\Phi \cdot \{x{:}\Psi(x)\} \subseteq \{x{:}\Phi(x)\}$
(4) $\lambda\Psi \cdot \lambda\Phi \cdot \{x{:}\Psi(x)\} = \{x{:}\Psi(x)\,\&\,\Phi(x)\}$

In (1–4), 'Ψ' and 'Φ' range over functions that map entities to truth values, with metalanguage sentences like '∀x:Ψ(x)[Φ(x)]' denoting truth values relative to assignments of values to variables. For example, if the semantic value of *groundhog* is λx·Groundhog(x)—a function of type <e, t>—then the semantic value of *every groundhog* is λΦ·∀x:Groundhog(x)[Φ(x)]; though this function of type <<e, t>, t> can also be specified in ways corresponding to (2–4). These familiar assumptions highlight a familiar question: can expressions with the same semantic value have different meanings? In this paper, we argue that they can.

To be sure, *meaning* is a vexed term that may not have an extension. But the question is whether theories of natural languages (i.e., the spoken or signed languages that humans naturally acquire) should employ semantic vocabulary that lets theorists distinguish extensionally equivalent expressions. As a case study, we consider a relation exhibited by the semantic values of predicates: the relation corresponding to universal quantifiers. Specifically, we focus on the distinction between specifying this relation in terms of individuals and a first-order quantifier as in (1) or (2), and specifying the relation in terms of sets or other devices for grouping the satisfiers of a predicate together as in (3) or (4). We argue that while *each* is first-order in the former sense, *every* and *all* are second-order in the latter sense.

If we think of meanings as mental representations that are encoded in particular formats, then it might seem obvious that two words for universal quantification can have distinct meanings, much like two names for the same planet (*Hesperus* and

[1] For these purposes, extensions include functions *in extension* from possible worlds to sets of entities, from variables to assignable values, from functions to functions, etc. Church (1941) contrasted these abstracta—identifiable, given sufficient ontology, with sets of ordered pairs—with what he called functions *in intension*, or *procedures* that have extensions, explicitly allowing that distinct procedures can have the same extension; cp. Frege's (1893) procedural notion of function, and Chomsky's (1986) notion of I-language.

Phosphorus) or two predicates used to talk about the members of a certain species (*woodchuck* and *whistlepig*). From this "internalist" perspective, *each* and *every* might have the same semantic value—the relevant function of type <<e, t>, <<e, t>, t>>—but nonetheless be nonsynonymous expressions.

But if we think of meanings as aspects of an environment (shared by speakers who mentally represent that environment idiosyncratically), then it might seem equally obvious that a word like *Venus* or *groundhog* has the same meaning for each competent speaker, regardless of how speakers represent the relevant extensions. From this "extensional/externalist" perspective, the fact that *each* and *every* share a semantic value implies that they have the same meaning, even if they differ in terms of their pronunciations and detailed syntactic features (cp. *likely* and *probable*).[2]

In this context, experimental investigation into how speakers understand logical and mathematical vocabulary can be especially useful. Unlike *Venus* or *groundhog*, it seems clear that a word like *every* or *four* corresponds to a family of concepts that are guaranteed to be extensionally equivalent across possible worlds. With this in mind, various authors have looked for symptoms of how speakers understand the quantificational determiner *most*, whose semantic value can be specified in many ways, including (5–8).

(5) $\lambda\Psi \cdot \lambda\Phi \cdot |\{x : \Psi(x) \,\&\, \Phi(x)\}| > 1/2|\{x : \Psi(x)\}|$
(6) $\lambda\Psi \cdot \lambda\Phi \cdot \mathrm{OneToOnePlus}(\{x : \Psi(x) \,\&\, \Phi(x)\},\ \{x : \Psi(x) \,\&\, \sim \Phi(x)\})$
(7) $\lambda\Psi \cdot \lambda\Phi \cdot |\{x : \Psi(x) \,\&\, \Phi(x)\}| > |\{x : \Psi(x) \,\&\, \sim \Phi(x)\}|$
(8) $\lambda\Psi \cdot \lambda\Phi \cdot |\{x : \Psi(x) \,\&\, \Phi(x)\}\ | > |\{x : \Psi(x)\}| - |\{x : \Psi(x) \,\&\, \Phi(x)\}|$

In particular, Halberda et al. (2008) show that understanding *most* is independent of knowledge of large exact number words. Hackl (2009) offers an argument against *most* being understood as *more than half*, as in (5). Pietroski et al. (2009) offer an argument against *most* being understood in terms of one-to-one correspondence as in (6); where a set **s** bears the OneToOnePlus relation to a set **s'** just in case the elements of some proper subset of **s** correspond one-to-one with the elements of **s'** (and, to accommodate infinite sets, not vice versa). Lidz et al. (2011) build on these results and offer evidence in favor of a specification like (8) over one like (7). Tomaszewicz (2011) shows that these same predictions are borne out for majority quantifiers in Polish. Knowlton et al. (2021) extend the predictions of (8) to a variety of tasks involving adults and children. Note that (5–8) are for combination with count nouns; see Odic et al. (2018) for details concerning mass nouns.

So while the meaning of *most* is somehow proportional and thus not first-orderizable (see Rescher, 1962; Wiggins, 1980; Barwise and Cooper, 1981), the relevant extension can be specified without invoking a proportion like ½. It can also

[2] Words like *unicorn*, *centaur*, and *ghost* present familiar difficulties for the extensional/externalist approach. These words are not true of anything, yet they have different meanings. Compare *picture of a centaur watching a ghost ride a unicorn* and *picture of a ghost watching a unicorn ride a centaur*; see, e.g., Goodman (1949, 1953), and Chomsky (1957). One response is to posit possible worlds at which unicorns exist (e.g., Lewis, 1986); though Kripke (1980) offered powerful reasons for not doing so. Alternatively, one can posit many "ways of presenting" the empty set; but this suggests that extensional semantic values are distinct from (even if determined by) lexical meanings.

be specified without appeal to cardinalities, or without appeal to any form of predicate negation. This made *most* ripe for psycholinguistic investigation. But *every*—or whatever word gets used to translate classical talk (in Greek or Latin) of universal quantification—presents similar issues, given the distinction between first-order and second-order quantifiers. And as Vendler's (1962) discussion made clear, English has several lexical devices for expressing (one or more concepts of) universal quantification. Each of these devices has its own quirks (see Sect. 1.2), raising the question of whether they have the same meaning.

1.1 First-order versus second-order representations

In one sense, (1–8) are all second-order specifications. They all describe relations exhibited by functions of type <e, t> as opposed to relations exhibited by individuals of type <e>. At least for present purposes, we assume that this is correct: *each*, *every*, and *all* have (categorematic) meanings that combine with the meanings of two predicates.[3] The issue is whether these words have a common meaning that can be identified with the higher-order relation/function/extension, or whether the words have distinct meanings that reflect distinct (and possibly first-order) specifications of a common extension. As an analogy, consider (9) and (10), two ways of specifying a function of type <<e, t>, <<e, t>, t>> corresponding to the numeric word *four*.

(9) $\lambda\Psi \cdot \lambda\Phi \cdot |\{x : \Psi(x) \,\&\, \Phi(x)\}| = 4$

(10) $$\begin{aligned}&\lambda\Psi \cdot \lambda\Phi \cdot \exists \mathrm{x} \exists \mathrm{y} \exists \mathrm{z} \exists \mathrm{w}\\ &\{\Psi(\mathrm{x}) \,\&\, \Psi(\mathrm{y}) \,\&\, \Psi(\mathrm{z}) \,\&\, \Psi(\mathrm{w}) \,\&\, \Phi(\mathrm{x}) \,\&\, \Phi(\mathrm{y}) \,\&\, \Phi(\mathrm{z}) \,\&\, \Phi(\mathrm{w}) \,\&\\ &(\mathrm{x} \neq \mathrm{y}) \,\&\, (\mathrm{x} \neq \mathrm{z}) \,\&\, (\mathrm{x} \neq \mathrm{w}) \,\&\, (\mathrm{y} \neq \mathrm{z}) \,\&\, (\mathrm{y} \neq \mathrm{w}) \,\&\, (\mathrm{z} \neq \mathrm{w}) \,\&\\ &\forall \mathrm{s} : \Psi(\mathrm{s}) \,\&\, \Phi(\mathrm{s})[(\mathrm{x} = \mathrm{s}) \vee (\mathrm{y} = \mathrm{s}) \vee (\mathrm{z} = \mathrm{s}) \vee (\mathrm{w} = \mathrm{s})]\}\end{aligned}$$

We can imagine minds that have the capacity to form both kinds of representations but find the first-orderized (10) too cumbersome to use. We can also imagine minds that are unable to form second-order representations like (9) but can form and use first-order representations like (10). If minds of both kinds can come to have words that are decently translated with *four*, it seems wrong for theorists to insist that both specifications of the function are equally good

[3] By contrast, assuming displacement of determiner phrases, we could adopt a syncategorematic treatment that posits sentence frames like $[_S\ [_D \text{ every } [_N \ldots]]_i\ [_S \ldots t_i \ldots]]$. Truth conditions for instances of this frame can be specified in a way that corresponds to (1): relative to any assignment A of values to indices, $\forall x{:}\ [\![[_N \ldots]]\!](x)[[\![[_S \ldots t_i \ldots]]\!]^{A:x/i}\ ]$; i.e., for each individual mapped to truth by the semantic value of the noun (or noun phrase), truth is the semantic value of the embedded sentence relative to the assignment that is just like A except that the individual in question is assigned to the index of the relevant trace of displacement. This way of specifying an extensional semantic role for $[_S\ [_D \text{ every } [_N \ldots]]_i\ [_S \ldots t_i \ldots]]$, which invites the further step of dispensing with appeals to truth values and functions as semantic values for predicates, is not without virtues; see Davidson (1967) and Higginbotham (1985). But our questions would remain, since the semantic role in question could also be specified in ways corresponding to (2–4).

specifications of how speakers understand those words. Put another way, even if we assume that the translations of *four* have the same extension, (9) and (10) may not be equally good as theoretical depictions of what the words mean.

Of course, in providing models of how lexical meanings can be combined, we often idealize away from many respects in which lexical meanings are complex. This is not only true for quantifiers like *every*, but for common nouns like *window* and *house*, proper nouns like *London* and *France*, and verbs like *bear* and *set*, all of which exhibit conceptual equivocality or “multi-dimensionality” (see Chomsky, 1977, 2000; Travis, 2008; Pietroski, 2018; a.o.). Extensional models of meanings that abstract away from the psychological details of how semantic values are mentally represented can be useful. But as Chomsky (1995, 2000) and others have stressed, details of human psychology have proven relevant to the study of natural language at many levels of analysis (e.g., Halle, 1978; Kazanina et al., 2006; Lidz and Musolino, 2002; Phillips, 2006). Decisions about which theoretical vocabulary to employ in formulating theories of natural language meaning, and about what such theories should explain, should likewise be made in light of all relevant evidence. If certain contrasts in how extensions are specified turn out to be explanatorily relevant, as suggested by the work on *most* cited above, this bolsters other reasons for developing an internalist conception of meaning.

We argue here that like the meaning of *most*, the meanings of *each*, *every*, and *all* are not neutral with respect to representational format. In particular, we think the distinction between first-order and second-order specifications is psychologically realized in a way that has behavioral repercussions. As mentioned above, the crucial contrast concerns whether or not the specification depends on some device for grouping the satisfiers of a predicate together. We don’t want to insist on any particular way of formalizing this distinction in advance. But to clarify the contrast, it may help to move away from functions—of type <e, t> for words like *groundhog*, type <<e, t>, t> for phrases like *every groundhog*, and type <<e, t>, <<e, t>, t>> for words like *every*—and consider theories that specify satisfaction conditions for lexical items, as in (11).

(11) a. for any entity x, $\text{Satisfies}(\text{x}, \textit{groundhog}) \equiv \text{x is a groundhog}$
b. for any entities x and y, $\text{Satisfies}(<\text{x},\ \text{y}>, \textit{above}) \equiv \text{x is above y}$
c. for any sets X and Y, $\text{Satisfies}(<\underline{\text{X,Y}}>, \textit{every}) \equiv \forall \text{x}:\text{x} \in \text{Y}[x \in X]$

In our sense, (11c) is a first-order specification. Even though it specifies a relation that is exhibited by sets—unlike the relation specified for *above* with (11b)—(11c) is described (on the right side of ‘≡’) in a way that doesn’t rely on representing the elements of either set collectively. By contrast, (12) is more like the essentially second-order (13); cp. (6) above.

(12) $\text{Satisfies}(<\underline{\text{X, Y}}>, \textit{every}) \equiv \text{Y} = \text{Y} \cap \text{X}$
(13) $\text{Satisfies}(<\underline{\text{X, Y}}>, \textit{most}) \equiv \text{OneToOnePlus}(\text{Y} \cap \text{X},\ \text{Y} - \text{X})$

While (12) and (13) are both second-order, only (12) is first-orderizable, in that the extension specified for *every* can also be specified with the first-order (11c). This is analogous to the specification of *four* in (9), which is not first-orderized but is first-orderizable (e.g., as (10)).

Likewise, the same first-order/second-order distinction arises if appeals to sets are replaced with plural quantification, as in Boolos (1984). On this approach, capitalized variables range over the individuals in some domain (that may or may not include sets), but one or more values can be assigned to each capitalized variable. In this way, we can specify the satisfiers of *every* as the ordered pairs of individuals that meet the following plural condition: their internal elements are the things that satisfy the relevant noun or noun phrase (e.g., *groundhog* in *every groundhog*); their external elements are the things that satisfy the other predicate (e.g., *is brown*); and each one of the internal elements is one of the external elements. This can be formalized as in (14); see Pietroski (2018).

(14) $\text{for any ordered pairs, the Os, Satisfies}(O, \mathit{every}) \equiv$
$\exists X \exists Y[\text{External}(O, X) \;\&\; \text{Internal}(O, Y) \;\&\; \forall z{:}Yz(Xz)]$

In our sense, '∀z:Yz(Xz)' is a first-order specification of how the internal elements of some ordered pairs must be related to the external elements in order to meet the condition imposed by *every*. By contrast, (15a) and (15b) are extensionally equivalent second-order specifications that could be substituted for '∀z:Yz(Xz)' in (14); where 'Include' allows for improper inclusion (a.k.a. identity), and (15b) exploits the fact that every Y is an X if and only if every Y is a Y that is an X.

(15) a. $\text{Include}(X, Y)$
b. $\exists Z[\text{Identical}(Z, Y) \;\&\; \forall x(Zx \equiv Yx \;\&\; Xx)]$

In short, even assuming that words for universal quantification have a shared semantic value, perhaps some of these words are understood as specifying that semantic value in first-order terms, while others are understood as specifying it in second-order terms. This invites a search for empirical evidence that distinguishes these two kinds of specifications. And if speakers do associate different words for universal quantification with these distinct representational formats, then for each such word, it is an empirical question whether a first-order or second-order specification better describes how speakers understand that word.

Our main prediction is that if speakers understand a quantifier *Q* in a second-order way, then given independently studied capacities for representing groups that humans share with many other animals, the phrase *Q big dot(s)* should facilitate representing the big dots in a way that supports encoding properties that depend on those group representations (e.g., cardinality). By contrast, we predict that if speakers understand *Q* in a first-order way, then *Q big dot(s)* should prompt them to represent each big dot as an individual, without promoting representation of the big dots as a group whose cardinality can be estimated. We test this by asking participants to evaluate a quantificational sentence like *Q big dot(s) is/are blue*, relative to a perceived scene, and then asking a follow-up question like "how many big dots were there?".

We find that participants have a better memory for cardinality following *all*-sentences than following otherwise equivalent *each*-sentences (experiment 1). We find a similar pattern of performance when comparing *every* and *each* in a separate group of participants (experiment 2). This second experiment is especially important as it provides a minimal pair: the truth-conditions, syntax, participants, and images are the same across conditions, with the only difference being the quantificational determiner. We find that pitting *each of the* against *all of the* yields similar though less conclusive results (experiment 3). Finally, we directly compare *every* with *all of the*; and here, we find an advantage for *all of the*, suggesting that partitivity and/or plurality is also involved in driving attention to groups (experiment 4).

One limitation of the current work is that we present four within-participant pair-wise comparisons. In the future, we plan to compare various quantificational expressions in a single between-subjects study. Still, taken together, we argue that these results provide evidence that *all* and *every* are represented in a second-order format, whereas *each* is represented as first-order. This difference (along with partitivity and plurality) encourages participants to use group-based or individual-based verification strategies, despite the truth-conditions being the same. But our claim is not that the first-order/second-order contrast exhausts the respects in which words for universal quantification can differ. There are a number of differences among the English universal quantifiers, outlined in the remainder of this section. Accounting for all of these differences will require further distinctions within the broad classes of first-order and second-order representations.

1.2 Semantic and grammatical properties of *each*, *every*, and *all*

Before turning to the experiments, it may be useful to review some independent reasons for suspecting that *each*, *every*, and *all* have distinct meanings. Some of these differences invite diagnosis in terms of a first-order/second-order contrast. However, they do not yet provide clear evidence of a corresponding psychological distinction: at least initially, one might bracket the grammatical differences across *each*, *every*, and *all* as quirks—or as Szabolcsi (2015) puts it, “annotations on the pertinent lexical items”—especially if one assumes that speakers regularly represent the same semantic value in different ways.

The universal quantifiers differ along several dimensions, including: ease of generic construals (Gil, 1992), speaker preferences regarding scope (Feiman and Snedeker, 2016; Ioup, 1975; Kurtzman and MacDonald, 1993), interactions with negation (Beghelli and Stowell, 1997), and compatibility with collective predicates (Dowty, 1987; Vendler, 1962). These results have led to a common impression that *each* directs attention to individuals, while *all* typically invites representations of groups.

For example, using *each* in (16) seems to imply many glancing-events, each targeting a different member of the flock, while using *all* conjures an image of a preacher looking out at his congregation with one prolonged stare (Beghelli and Stowell, 1997).

(16) The preacher looked at {each/all} of the members of his flock.

This is in line with the observation discussed by Vendler (1962) that *each* forces distributivity and is therefore incompatible with collective predicates, as in (17a). On the other hand, *all* is easily used with many collective predicates; though it doesn't force collective interpretations, as illustrated with the distributive (17b).[4]

(17) a. {*Each/All} of the soldiers surrounded the fortress.
b. {Each/All} of the soldiers admired the fortress.

Grammatically, *each* requires the partitive *of* when used with a plural noun phrase like *the dogs* in (18a), whereas *all* can take the plural noun phrase *the dogs* or the full prepositional phrase *of the dogs*.

(18) a. Each *(of) the dogs barked.
b. All (of) the dogs barked.

Though in *all the dogs*, it may be that *all* intensifies *the*, as opposed to serving as a quantificational determiner. As Baker (1995) puts it, it might be that "*all* is not quantificational; it only emphasizes that the entire referent picked out by *the NP* is involved."

It may be more significant that *one* can be added to sentences with *each*, without a change in meaning, as in (19a). But with *one*, (19b) sounds defective; though given the right context and prosody, it can be used to convey that the plurality presupposition of *all* is violated (e.g., in snarky response to someone who asked if all the dogs barked, in a context that includes only one dog).

(19) a. Each (one) of the dogs barked.
b. All (*one) of the dogs barked.

Every tends to pattern with *each* rather than *all*. This is especially clear with regard to compatibility with collective predicates, as shown in (20); see Vendler (1962), Dowty (1987).[5]

(20) a. {*Each/?Every} dot is alike.
b. All dots are alike.

Likewise, *every* requires a singular count noun, as in (21a). But it differs from *each* and *all* in being incompatible with the partitive without support by *one*, as seen in (21b).

[4] And in (17a), *the* matters; cp. *All soldiers {*surrounded/admired} the fortress*. Some collective predicates, like *be numerous* and *be a good team*, cannot be used with any of the universal quantifiers (Champollion, 2015; Dowty, 1987; Winter, 2002). But since these predicates also conflict with other quantifiers, including the manifestly second-order *most*, this is not a diagnostic for second-order quantification.

[5] Moreover, as Vendler notes, *All of those dots are similar* can be heard as true even if there is no single dimension on which each pair of dots is similar. But it's much harder to find a sensible interpretation for *Each of those dots is similar* or *Every dot is similar*. There are also cases where sentences with *each* can give rise to distributive readings, but those with *every* cannot; see Sect. 5.

(21) a. Every {dog is/*dogs are} brown.
b. Every*(one) of the dogs barked.

There are, however, a few respects in which *every* patterns with *all*. Beghelli and Stowell (1997) note that both words can occur with *almost*—as if they both indicate the end point of a scale—and with negation. In these respects, *each* is the odd universal quantifier out, as shown in (22) and (23).

(22) a. One kid ate almost {all the cookies/every cookie}
b. *One kid ate almost each cookie
(23) a. Not {all the kids/every kid} ate a cookie
b. *Not each kid ate a cookie

Perhaps relatedly, *every* is friendly to generic interpretations in a way that *each* is not. For example, even though (24a) seems like a distributive generalization that is truth-conditionally equivalent to (24b), (24c) carries no generic implication of the sort invited by (24a) or (24b). It's worth noting, though, that *all* only has generic import when it takes a bare NP, as in (24b). When it occurs with a definite plural or partitive, as in (24d), it carries no such generic implication.

(24) a. Every rabbit hops.
b. All rabbits hop.
c. Each rabbit hops.
d. All (of) the rabbits hop.

These facts suggest differences in the syntax and semantics of *each*, *every*, and *all*. A theme throughout these examples is that sentences with *each* seem to be about individuals in some sense, and that this is less true for sentences with *all* (and perhaps for sentences with *every*). Indeed, this suspicion has been discussed in the literature that treats *all* as quantifying over pluralities of some sort (e.g., Champollion, 2015; Dowty, 1987; Križ, 2015; Winter, 2002). Here, we offer independent psychological evidence to this effect. It is outside the scope of this paper, though, to adjudicate between the various alternative theories that attempt to capture the above differences or relate them to linguistic differences in other domains. Our aim is more modest and more general: to provide evidence of the first-order/second-order representational distinction by testing the simple linking hypothesis that it suggests.

2 Motivating a psychophysical investigation of quantifier meanings

Evidence for differences in representational format can come from various sources, including studies of how humans verify quantificational sentences in controlled contexts. Lidz et al. (2011) argue that all else equal, people are biased to evaluate a given sentence with a procedure that transparently reflects that sentence's meaning.[6]

[6] Their claim is not that meanings *are* verification strategies or that people *always* use a certain strategy to evaluate a given sentence. In ordinary (uncontrolled) contexts, many considerations can be relevant to choosing a verification strategy. The claim is rather that the representational format of the meaning

The idea, in the spirit of Marr (1982), is that the representational format of a contentful thought can highlight the applicability of certain computational procedures, thereby making some ways of evaluating that thought more natural than others; see also Pietroski et al. (2011).

As an example, Pietroski et al. (2009) show that participants verify *most*-sentences with a cardinality-based strategy even when superior alternatives are readily available. More specifically, when asked whether *most of the dots are blue,* participants showed clear signatures of using the Approximate Number System; for discussion of this system, see Feigenson et al. (2004); Dehaene (2011). Participants did so even for displays where using the Approximate Number System was a sub-optimal strategy. When given identical displays but prompted to use a one-to-one correspondence strategy (with a question not involving *most*), participants not only did so, they were faster and more accurate than when relying on cardinality comparisons. This bolstered the case that the meaning of *most* is specified in terms of cardinality, not correspondence.

As discussed in the introduction, we predict that the distinction between first-order and second-order representations should similarly contribute to determining what verification strategy is used. If a given quantifier has a first-order representation, then evaluating sentences with it should, all else equal, lead participants to represent individuals as such (perhaps verifying the sentence by checking for disconfirming individuals). If a given quantifier has a second-order representation, then evaluating sentences with that quantifier should, all else equal, lead participants to represent groups as such (perhaps verifying the sentence based on the level of variance within the group).

One attraction of testing this prediction on universal quantifiers is that the sentences being compared at least approximate grammatically minimal pairs—e.g., *each/every big dot is blue*. But we need an independent measure of whether participants are representing a group or representing individuals as such. Fortunately, it is well-established in Cognitive Psychology that these two modes of representing objects—as ensembles or as individuals—have distinct behavioral signatures, observable as early as infancy (for a short review of the cognitive systems for representing ensembles and object-files, see Feigenson et al., 2004; for a longer review, see Carey, 2009).

One consequence of attending to and representing a group is that knowledge of its summary statistics—center of mass, density, average size, approximate cardinality, etc.—becomes available (e.g., Ariely, 2001; Chong and Treisman, 2003; Halberda et al., 2006; Burr and Ross, 2008; Alvarez, 2011; Whitney and Yamanashi Leib 2018). Representing a group and encoding knowledge of its summary statistics does not even require explicitly representing each individual constituting that group (Alvarez and Oliva, 2008; Ariely, 2001). On the other hand, merely attending to and representing individuals plausibly will not afford the same

Footnote 6 continued
contributes a detectable influence over verification procedures. And if other considerations are controlled for, then variation in verification strategies for different expressions can reasonably be attributed to variation in the way the meanings of those expressions are mentally encoded.

access to summary statistics of any groups those individuals belong to. That is, participants may begin to form some sense of summary statistics of, for example, the big dots as they attend individual big dots, but this sense should not be as immediate or accurate as attending the group of big dots as such. In this paper, we focus on one ensemble summary statistic: cardinality.

2.1 Measuring cardinality knowledge

There are two things to consider when judging the extent of someone's cardinality knowledge across many trials. The first is accuracy: on average, how much do they overestimate or underestimate (Stevens, 1964)? The second is precision: how reliable—as measured by standard deviation from the mean—are their responses (Laming, 1997)? These two parameters are captured by the standard psychophysical model of magnitude estimation: Gaussian tuning curves ordered along an internal scale (Fig. 1)

According to this standard model, perceived numerosities "activate" different parts of the scale, here called the "mental number line".[7] These distributions are linearly ordered, with larger numerosities represented farther to the right. In Fig. 1, each Gaussian is centered over the corresponding numerosity on the mental number line (i.e., this hypothetical participant is perfectly accurate). If the participant modeled by Fig. 1 were shown an image of say, 9 dots on a computer screen, and asked to estimate (without counting) how many there were, they would answer "nine" most often, sometimes answering "eight" or "ten", less frequently "seven" or "eleven", and so on (i.e., they are not perfectly precise).

In an actual case, the relevant signal from the environment might be compressed in the process of building a representation of the correspondng numerosity, leading to underestimation. Informally, for a particular perceiver, the distribution of activation that arises from seeing, say, 13 items in the world might systematically result in Gaussian activation centered over 11 on their mental number line. Such a participant will systematically underestimate if shown 13 dots, usually answering "eleven", sometimes answering "ten" or "twelve", and so on. This tendency to underestimate is known as *inaccuracy*, and is captured by the parameter β in (25), where y is the percept (i.e., the participant's numerical estimate, on average), x is the actual number of objects presented, and α is a scaling factor.[8]

(25) $y = \alpha x^{\beta}$

Figure 1 represents the ideal case, where $\beta = 1$. Intuitively, this means that every distribution is centered over the "correct" value on the mental number line, so this

[7] This model does not only apply to numerosities perceived visually. It has been applied to many other psychological dimensions as well (e.g., loudness, brightness, distance) across multiple modalities (e.g., vision, audition, touch) (Cantlon et al., 2009; Lu and Dosher, 2014; Odic et al., 2016; Stevens, 1964).

[8] The parameter β in (25–26) can either be thought of as representing the degree of compression/expansion of the signal or of the response code that the Gaussian distributions of Fig. 1 are mapped to (see Izard and Dehaene, 2008).

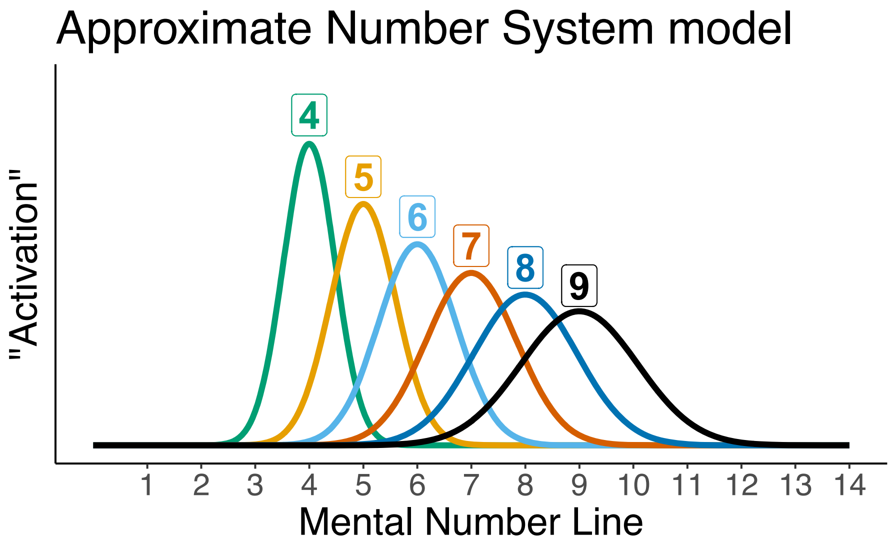

Fig. 1 Gaussian tuning curves along an internal "mental number line"

participant will respond correctly, on average, when shown any numerosity. In practice, adults have a $\beta < 1$ for cardinality estimation, meaning we systematically underestimate, with the amount of underestimation becoming more drastic as the represented numerosity increases (e.g., Krueger, 1984; Odic et al., 2016).

Another one of the model's hallmarks is scalar variability: the fact that the standard deviation of the Gaussian "activation patterns" increases linearly with the mean. The distribution centered on 9 in Fig. 1, for instance, is much wider than the distribution centered on 4. This captures the fact that representing larger numerosities is a "noisier" process than representing smaller numerosities (with a corresponding decrease in confidence; see Halberda and Odic, 2015). Scalar variability gives rise to the well-documented ratio-dependence that comes with numerosity estimation (e.g., 9 and 10 are just as difficult to distinguish as 90 and 100). It also explains why one is more confident about answering "five" after seeing 5 dots than about answering "fifty" after seeing 50: the distribution activated when experiencing the numerosity 50 has a larger standard deviation and thus overlaps with more numbers on the mental number line.

The rate of increase in standard deviation as number increases—which we'll call σ—develops throughout the lifespan (Halberda and Feigenson, 2008) and is subject to individual differences (Halberda et al., 2012; Libertus et al., 2012). If a participant has a large σ, their estimates will be more variable, and that variability will grow rapidly as numerosities increase. If a participant has a small σ, their estimates will have less variability, and that variability will grow at a slower rate as numerosities increase.

As an example, consider two hypothetical participants. If both have perfect accuracy ($\beta = 1$) and are shown fifteen dots, then both will respond "fifteen" on average. But suppose they have different precisions. If the first participant has a σ of .2 and is shown fifteen dots, they will respond "fifteen" about 13% of the time. If the second participant has a σ of .4 and is shown the same fifteen dots, they will respond "fifteen" only about 7% of the time. Likewise, if shown thirty dots, both participants will respond "thirty" on average. But the highly precise participant will respond "thirty" about 7% of the time whereas the less precise participant will respond "thirty" only about 3% of the time.

In sum, there are two main measures of cardinality estimation ability: accuracy (β) and variability (σ). Responses (*y*) to being shown some numerosity (*x*) are thus modeled as the Gaussian distribution in (26).

(26) $y \sim N\left(\alpha x^{\beta}, \sigma \times \alpha x^{\beta}\right)$

As mentioned, the important point for our task is that attending to and representing a group as such enhances sensitivity to that group's cardinality, compared to attending to and representing the individuals constituting that group. In the above terms, the enhanced sensitivity to numerosity should correspond to an increase in β and a decrease in σ.

Importantly, by using β and σ as our dependent measures, we do not mean to suggest that participants' *internal* accuracy and precision changes from situation to situation. Our understanding of the psychophysical model described above is that accuracy and precision are constant within an individual at a given time. But when there is no attention to the objects queried, participants are unable to build an internal representation of their cardinality. This will result in guessing some proportion of the time. These guesses—which will probably involve a good deal of reasoning and strategizing—have the effect of decreasing the *observed* accuracy and precision.

In each of our experiments, participants were shown dot-displays, like Fig. 2, and asked to evaluate sentences like *all big dots are blue*. After answering TRUE or FALSE, the image disappeared, and participants were asked to recall the cardinality of some group of dots from the display. Sometimes they were asked about a distractor set, like "how many small dots were there?". Because the set of small dots is irrelevant to the sentence *all big dots are blue*, these distractor questions serve as a baseline measure of participants' observed accuracy and precision in the absence of attention to the relevant group.

Other follow-up questions probed the target set (i.e., the one denoted by the quantifier's internal argument). In the *all big dots are blue* example, the target question would be "how many big dots were there?". If *all* has a second-order meaning, we expect participants to be biased to attend to the set of big dots when evaluating the sentence. In doing so, they should have a good estimate of its cardinality (i.e., a higher observed β and lower observed σ compared to the distractor baseline). If *all* has a first-order meaning, we instead expect them to consider individual dots. In using this strategy, they should have an estimate of the big dots closer to the distractor baseline. Of course, regardless of details about the quantifier's mental representation, performance should be better when asked about the target set than when asked about the distractor set (if only for the fact that some

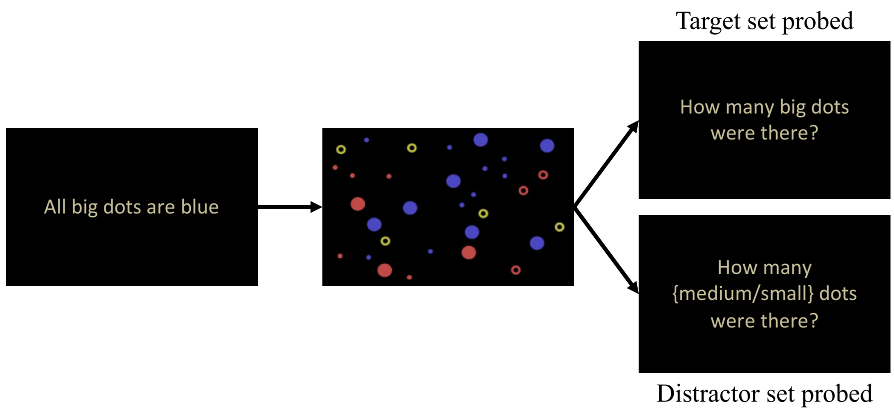

Fig. 2 The trial structure of the experiments

attention will be driven to the big dots by mentioning them). Our main diagnostic then, will be this: how well do participants know the cardinality of the relevant group after evaluating some quantificational sentence compared to after evaluating a truth-conditionally equivalent quantificational sentence?

To foreshadow our results, we find that sentences with *every* and *all* lead participants to have better estimates of the relevant sets' cardinalities relative to sentences with *each*. Less surprisingly, we also find a role for plurality. We find these results despite the fact that the truth-conditions, images, and participants are held constant. Given that there is no other reason to switch strategies between *every*-sentences and *each*-sentences, we argue these results reflect a difference in meaning; specifically, the proposed first-order/second-order distinction.

3 Experiment 1: *each* vs. *all*

3.1 Method

3.1.1 *Subjects*

27 University of Maryland undergraduates participated for course credit. All were native speakers of English. In this and subsequent experiments, we removed any participants who (a) scored under 85% accuracy on the TRUE/FALSE portion of the task, (b) reported using an explicit counting strategy, or (c) failed to complete the experiment. In this case, three participants were excluded from further analysis for being unable to complete the experiment in the allotted hour. This left us with 24 participants.

3.1.2 Materials

Sentences shown to participants had one of two forms: *each [size] dot is [color]* or *all [size] dots are [color]*. Dot displays consisted of a black background with red, yellow, and blue dots that could be big, medium, or small (see Fig. 2 above). Medium dots had black holes in the middle, to make them more distinguishable from the other two sizes (Chen, 1982, 2005). The dot sizes and colors were named during the training portion of the experiment to ensure that they could be correctly identified by participants.

Each display consisted of between 24 and 48 dots, with six size/color combinations (e.g., big blue dots) present. This was to ensure that participants could not enumerate each subset to prepare for all possible follow-up cardinality questions, as participants can only enumerate two subsets and the superset in parallel (Halberda et al., 2006). Each size/color combination that was present contained a minimum of 3 and a maximum of 9 dots.

Cardinality questions were distributed as follows: 30 questions probed the target size (e.g., big dots, in the example above), 30 probed a distractor size (e.g., small or medium dots), 30 probed the target color (e.g., blue dots), 30 probed a distractor color (e.g., red or yellow dots), and 16 probed the total number of dots. We did not ask participants about any subsets defined by the conjunction of two predicates (e.g., big blue dots), though see Knowlton et al. (2020), which used a similar method and probed size/color combinations.

Here, we report and analyze only the results following target and distractor size trials (the others were included as filler trials to ensure that participants could not guess in advance which group they would be asked about). This and subsequent experiments were built and run using PsychoPy2 (Peirce et al., 2019) and all data analysis was carried out in R (R core team, 2017).

3.1.3 Procedure

Participants completed 272 trials in a 2 (quantifier) 2 (set probed) design. On each trial, they read a quantificational sentence. Type of quantificational sentence was blocked and initial condition was counterbalanced such that half of the participants started with *each*-sentences and half of the participants started with *all*-sentences. Participants were allowed a short break between blocks after which the instructions for the second block were given verbally (instructions for both blocks were identical save for the quantifier used).

After reading the initial sentence and pressing ‘space’, participants viewed a dot display and evaluated whether that sentence was TRUE or FALSE with respect to the display by pressing ‘J’ (for TRUE) or ‘F’ (for FALSE). Dot displays remained on screen until participants offered their judgment (though they were told to respond as quickly as possible, to discourage explicit counting). Then, the screen went blank and they were asked to give a cardinality estimate of one of the subsets present in the display by typing in a number and pressing ‘enter’.

3.2 Predictions

Given some of the linguistic facts reviewed in Sect. 1.2, *each* is the most likely of the universals to have a first-order representation and *all* is the most likely universal to be specified in a second-order way. If this is right, we should expect better accuracy and precision on target questions following *all*-sentences than on the same questions following *each*-sentences. Moreover, *all* combines with a plural noun phrase (e.g., *big dots*), whereas *each* combines with the singular (e.g., *big dot*), which may also modulate attention to groups.

Regardless, distractor questions are predicted to lead to worse overall performance than target questions. Participants have no reason to know, for example, the number of small dots after being asked whether *all big dots are blue*. For that reason, distractor questions serve as a kind of baseline measure of poorest possible cardinality estimation ability. If *all* is mentally represented in second-order terms, evaluating sentences with *all* should lead to more improvement over this baseline than evaluating truth-conditionally equivalent sentences with *each*.

Figure 3 provides a visual depiction of these predictions (in terms of percent error) for this and subsequent experiments.

3.3 Results and discussion

On the TRUE/FALSE portion of the task, participants correctly evaluated 96.7% of the *all*-sentences and 96.4% of the *each*-sentences. Figure 3 provides a visualization of participants' performance on the follow-up cardinality questions in terms of average percent error (this figure includes data from the current experiment and the three subsequent experiments). Lower percent error indicates better performance.

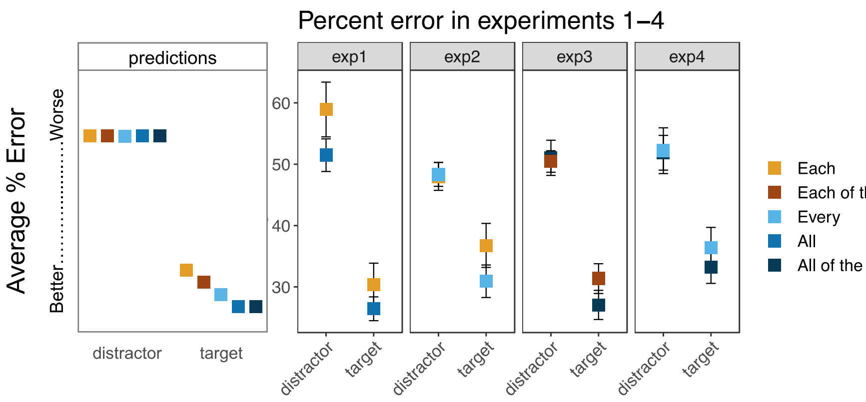

Fig. 3 Left: predictions if cardinality estimation ability is driven by quantifier and plurality, and if *each* is first-order while *all* and *every* are second-order. Right: percent error in each condition of experiments 1-4 (Exp1: *All*/*Each*; Exp2: *Each*/*Every*; Exp3: *Each of the*/*All of the*; Exp4: *All of the*/*Every*). Higher percent error reflects poorer estimations. For example, if the actual number of dots shown was 10, a response of 8 or 12 would result in 20% error; a response of 6 or 14 would result in 40% error

To analyze the responses on follow-up cardinality questions, we fit all of the data to different versions of the cardinality estimation model in (26). The "set probed only" model, (27a), allows the values for accuracy and precision to vary only based on the set probed (target or distractor). The "main effects" model, (27b), allows accuracy and precision to vary based on the set probed and the quantifier (in this case, *each* or *all*). The "interaction" model, (27c), includes terms for a main effect of set probed and terms for the interaction between quantifier and set probed. Finally, the "full" model, (27d), includes terms for both main effects and their interaction. In all models in (27), y is the participants' numerical response, x is the actual number of dots they were shown, Q is the quantifier used, S is the set that was probed in the follow-up question, β is accuracy, σ is variability, and α is a scaling factor.

The set probed only model in (27a) corresponds to the null hypothesis that the quantifier used has no effect on participants' cardinality estimates. The other three models correspond to three different ways of spelling out the alternative hypothesis that the quantifier does have an impact on participants' cardinality estimation ability.

(27) a. Set probed only:

$$y \sim N\begin{pmatrix} mean = (\alpha_0 + \alpha_1 S)x^{\beta_0+\beta_1 S} \\ sd = (\sigma_0 + \sigma_1 S) \times mean \end{pmatrix}$$

b. Main effects:

$$y \sim N\begin{pmatrix} mean = (\alpha_0 + \alpha_1 S + \alpha_2 Q)x^{\beta_0+\beta_1 S+\beta_2 Q} \\ sd = (\sigma_0 + \sigma_1 S + \sigma_2 Q) \times mean \end{pmatrix}$$

c. Interaction:

$$y \sim N\begin{pmatrix} mean = (\alpha_0 + \alpha_1 S + \alpha_3 SQ)x^{\beta_0+\beta_1 S+\beta_3 SQ} \\ sd = (\sigma_0 + \sigma_1 S + \sigma_3 SQ) \times mean \end{pmatrix}$$

d. Full:

$$y \sim N\begin{pmatrix} mean = (\alpha_0 + \alpha_1 S + \alpha_2 Q + \alpha_3 SQ)x^{\beta_0+\beta_1 S+\beta_2 Q+\beta_3 SQ} \\ sd = (\sigma_0 + \sigma_1 S + \sigma_2 Q + \sigma_3 SQ) \times mean \end{pmatrix}$$

The interaction model in (27c) is most in line with our predictions, as outlined above. That is, *all* should lead to group-representation and thus better cardinality estimation than *each* when a target set is probed, but we do not predict the choice of quantifier to have an effect on distractor questions (e.g., "how many small dots?" after evaluating *all big dots are blue*).

All models were fit using maximum likelihood estimation (the analysis script and data are available here: https://osf.io/a7y9w/). To compare the models, we considered both the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values, which reward models for capturing the data and penalize

Table 1 Model comparisons for experiment 1

Model	AIC	BIC
Set probed only	15861.37	15897.07
Main effects	15802.18	15855.72
Interaction	15813.46	15867.00
Full	**15757.98**	**15829.38**

Best model comparison values in bold

them for including greater numbers of parameters (Akaike, 1973; Schwarz, 1978; Stone, 1979). Lower values are indicative of striking a better trade-off between fit and complexity. As seen in Table 1, the Full model fares best according to both measures of model fit.

This strongly suggests that the choice of quantifier matters for cardinality estimation. In particular, sentences with *all* lead to better number-knowledge than sentences with *each* despite these sentences giving rise to the same answers upon evaluation. This result is well explained if *each* has a first-order representation that encourages participants to adopt an individual-based strategy, whereas *all* has a second-order representation that encourages them to represent groups. That said, there is also the possibility that the plural agreement on *dots* in the *all*-sentences drives the attention to groups that results in enhanced cardinality knowledge (i.e., the *each*-sentences and *all*-sentences are not true minimal pairs). Experiment 2 addresses this shortcoming by comparing *each* and *every*, which both require singular agreement, and finding an advantage for *every*.

It is surprising that, in the current experiment, the Full model provides a better fit than the Interaction model. This means the benefit afforded by *all* is present even on distractor questions, in which the set probed was not the set mentioned in the initial sentence that was evaluated. We suspect this aspect of the result is spurious, as it does not replicate in subsequent experiments.

4 Experiment 2: *each* vs. *every*

4.1 Method

4.1.1 Subjects

30 University of Maryland undergraduates participated in exchange for course credit. All were native speakers of English. Two participants were excluded for mentioning during debriefing that they used an explicit counting strategy and four were excluded for failing to finish both blocks in the allotted hour. This left us with 24 participants.

4.1.2 Materials

Sentences had one of two forms: *each [size] dot is [color]* or *every [size] dot is [color]*. As noted earlier, these sentences are true minimal pairs as only the quantifier differs between them. Aside from the sentences, the materials were identical to those used in experiment 1.

4.1.3 Procedure

The procedure of experiment 2 was identical to that of experiment 1.

4.2 Predictions

If *every* has a second-order representation, as experiment 1 suggests that *all* does, we should expect better accuracy and better precision on target questions following *every*-sentences than following *each*-sentences. If *every* instead has a first-order representation or if it was solely the difference in plural marking that drove the results of experiment 1, then we should expect the quantifier to make no difference to accuracy or precision on target questions. In any case, we predict that the quantifier should make no difference in accuracy or precision on distractor questions (a prediction that was surprisingly not supported in experiment 1) and that distractor questions will lead to worse performance than target questions (a prediction that was supported in experiment 1).

4.3 Results and discussion

On the TRUE/FALSE portion, participants correctly evaluated 96.9% of the *each*-sentences and 97.2% of the *every*-sentences. On the cardinality question, participants' performance bears out the predictions of *every* having a second-order representation. Table 2 presents the model comparison values. Both the AIC and BIC values favor the predicted interaction model.

Experiment 2 shows that *every* patterns like *all* in that it encourages group representation to a greater extent than *each*. However, experiment 2 offers stronger evidence than experiment 1 for two reasons. First, in experiment 2 the sentences form a minimal pair: since plurality was not a factor, the only explanation for the difference in cardinality estimation ability on target trials is the change in quantifier.

Table 2 Model comparisons for experiment 2

Model	AIC	BIC
Set probed only	15687.75	15723.48
Main effects	15653.59	15707.18
Interaction	**15636.82**	**15690.41**
Full	15640.33	15711.79

Best model comparison values in bold

Second, experiment 2 shows that the Interaction model is the best fit according to both ways of balancing reward for fit and penalty for parameters. This suggests that the choice of quantifier matters in exactly the predicted way: sentences with *every* lead to better number-knowledge than sentences with *each* only on questions probing the target set (i.e., the one that was mentioned in the initial statement).

In sum, experiment 2 (which tests a minimal pair) further supports the conclusion of experiment 1 that *each* has a first-order representation and thus dissuades participants from representing the relevant dots as a group. In addition, it suggests that *every*, like *all*, has a second-order representation that encourages representation of its first argument as a group. Experiments 3 and 4 further explore the role of plurality and partitivity in encouraging participants to represent the relevant dots as a group as opposed to independent individuals.

5 Experiment 3: *each of the* vs. *all of the*

5.1 Method

5.1.1 Subjects

30 University of Maryland undergraduates participated for course credit. All were native speakers of English. One was excluded from further analysis for scoring below 85% on the TRUE/FALSE portion, and five were excluded for being unable to complete the experiment in the allotted hour. This left us with 24 participants.

5.1.2 Materials

Sentences had one of two forms: *each of the [size] dots are [color]* or *all of the [size] dots are [color]*. As in experiment 2, these sentences constitute minimal pairs, so any differences in cardinality estimation ability can be more confidently blamed on the different quantifiers. Aside from the sentences, the materials were identical to those used in experiments 1 and 2.

5.1.3 Procedure

The procedure of experiment 3 was identical to that of experiments 1 and 2.

5.2 Predictions

Experiment 1 showed that *all*-sentences lead to better performance on cardinality questions than *each*-sentences. Experiment 2 controlled for the singular/plural distinction and found a similar difference between *every*-sentences and *each*-sentences, so it is unlikely that the singular/plural distinction alone accounted for the entirety of the difference found in experiment 1. Still, the singular/plural distinction and perhaps the partitive (*of the*) likely plays some role in driving attention to

Table 3 Model comparisons for experiment 3

Model	AIC	BIC
Set probed only	15344.01	**15379.72**
Main effects	15345.98	15399.55
Interaction	**15337.37**	15390.94
Full	15340.89	15412.32

Best model comparison values in bold

individuals or groups. To the extent that they do, the effect should disappear when comparing *each of the* and *all of the*. On the other hand, to the extent that the quantifier alone is responsible for encouraging participants to represent the relevant groups, then we should observe results similar to those in experiment 1.

5.3 Results and discussion

On the TRUE/FALSE portion, participants correctly evaluated 96.6% of the *all of the*-sentences and 96.1% of the *each of the*-sentences. Results from the cardinality portion were less conclusive. As seen in Table 3, the AIC and BIC values favor different models, owing to the different ways of balancing reward for good fit and penalty for achieving fit via parameters (BIC penalizes number of predictors more than AIC). The AIC value favors the predicted Interaction model, but the BIC value favors the Set probed only model.

Absent reason for preferring AIC over BIC, or vice versa, this result is inconclusive by itself. If either quantifier leads to better cardinality knowledge, it is *all* as opposed to *each*. But there is also reason to prefer the null hypothesis that quantifier plays no role.

Given the results of experiments 1–3 taken together, it seems that while *each* does not drive attention to groups to the same extent as *all* or *every*, participants are more likely to adopt a group-based verification strategy when *each* appears with the partitive *of the dots*. This would not be surprising, since *the dots* is a plural noun phrase that may itself encourage attention to groups. From this perspective, the interesting finding from experiment three would be that combining *each* with a partitive that encourages attention to groups leads to slightly worse performance than combining *all* with a matched partitive, as if *each* "undoes" some of the attention to groups that was encouraged by the plural. To further explore the role of the partitive/plural, experiment 4 directly tests its influence over and above the influence of the quantifier.

6 Experiment 4: *every* vs. *all of the*

6.1 Method

6.1.1 Subjects

28 University of Maryland undergraduates participated for course credit. All were native speakers of English. Four participants were excluded for being below 85% accuracy on the TRUE/FALSE portion. This left us with 24 participants.

6.1.2 Materials

Sentences had one of two forms: *every [size] dots is [color]* or *all of the [size] dots are [color]*. Aside from the sentences, the materials were identical to those used in experiments 1–3.

6.1.3 Procedure

The procedure of experiment 4 was identical to that of experiments 1–3.

6.2 Predictions

Given the results of experiment 3, the addition of partitive and/or plurality seems to play a role in encouraging participants to adopt a group-based strategy. But it remains to be seen whether this can improve participants' cardinality knowledge over and above the improvement they see from evaluating a sentence with a second-order quantifier like *every* or *all*. If it can, we expect *all of the*-sentences to lead to better performance than *every*-sentences. If participants are already at ceiling performance when evaluating a second-order quantifier, then both sentences should lead to the same level of cardinality knowledge.

6.3 Results and discussion

On the TRUE/FALSE portion, participants correctly evaluated 95.7% of the *all of the*-sentences and 95.6% of the *every*-sentences. Figure 3 shows that there is a slight

Table 4 Model comparisons for experiment 4

Model	AIC	BIC
Set probed only	16059.66	16095.35
Main effects	16048.99	16102.53
Interaction	**16027.27**	**16080.81**
Full	16030.65	16102.03

Best model comparison values in bold

accuracy improvement on target questions following *all of the*-sentences compared to *every*-sentences. Indeed, both the AIC and BIC values in Table 4 favor the Interaction model.

In sum, experiment 4 demonstrates that partitivity and plurality can effect participants' cardinality judgments over and above the quantifier. Experiments 1–3 suggested that both *every* and *all* have second-order formats, but nonetheless, experiment 4 shows that evaluating a sentence that includes *all of the big dots* leads to better performance than evaluating the corresponding sentence with *every big dot*.

7 General discussion

We began with the question of whether expressions that share a same semantic value can differ in meaning. As a case study, we considered the universal quantifiers, which might be specified in first-order or second-order representational formats. This logical distinction corresponds to a psychological distinction between individual ("object-file") and group ("ensemble") representations. We leveraged this connection in looking for evidence regarding the nature of the representations underlying *each*, *every*, and *all*. In particular, we probed participants' memory for various groups (by asking how well they could estimate the corresponding cardinalities) following evaluation of different quantificational sentences. The central idea is that a second-order quantifier should bias participants to attend to and represent groups during evaluation, in turn resulting in more accurate and more precise cardinality estimates for the relevant group (e.g., the big dots in *all of the big dots*). A first-order quantifier, on the other hand, should invite attending to and representing individuals, and in turn result in worse accuracy and precision when a participant is asked to estimate the cardinality of a set composed of those individuals.

Our main findings come from the first two experiments: sentences with *each* led to worse cardinality estimates for the relevant sets than truth-conditionally equivalent sentences with *all* or *every*. Less surprisingly, there also seems to be a role for partitivity and plurality, as the effect is reduced when *each dot* is replaced by *each of the dots* and, when compared head-to-head, *all of the*-sentences led to better cardinality knowledge than *every*-sentences. The resulting picture is one in which participants can evaluate a universally quantified sentence through the use of two different verification strategies: one rooted in attending to and representing individuals and the other rooted in attending to and representing groups. Which strategy they use in a given case is influenced by whether the first argument is partitive/plural or singular and by whether the quantifier is represented in a first-order or second-order format. Methodologically, detecting the latter difference is easier to the extent that the expressions compared are minimal pairs (i.e., experiment 2 offers more decisive evidence than experiments 1 or 3).

To be sure, there are other possible explanations besides the proposed first-order/second-order distinction. The remainder of this section considers some such alternative explanations.

7.1 Could this be about distributivity?

As we saw in Sect. 1, sentences with *each* are always given distributive interpretations such that the predicate in an *each*-sentence applies to the individuals in the domain (e.g., (28a) is not a good description of a situation in which the students sang together). On the other hand, *all* is not always used in this way. The judgment generally reported in the literature (e.g., Beghelli and Stowell, 1997) is that the ease of accepting the distributive interpretation (in which there were as many singings as there were students) is *each* > *every* > *all*, as in (28).

(28) a. Each student sang happy birthday (by themself/# in perfect harmony).
b. Every student sang happy birthday (by themself/? in perfect harmony).
c. All the students sang happy birthday (by themselves/in perfect harmony).

The linking hypothesis between this apparent fact and our experimental task might seem to be the following: because *each* is always used with distributive predicates, and because it is generally peoples' first choice of universal when expressing a distributive thought, people are naturally biased to think about individuals after hearing/reading *each*.

Notice that this linking hypothesis is about the word *each* being highly associated with distributive thoughts, not about the sentences in question being or not being distributive. All of the sentences used in our task contained stative predicates like *be blue* and there is no non-distributive way for *all of the dots* to be blue. Even when groups are implicated (as we suspect they are for *every* and *all*) the property of blueness is distributed down to the individual members. This raises a potential problem for an account of the results based only on distributivity: if the distributivity associated with *each* leads participants to represent individuals, why does the distributivity of the predicate *be blue* not have the same effect in a sentence like *every dot is blue*?

Another potential problem with a proposal based purely on distributivity is that *every* largely patterns like *each* with respect to giving rise to distributive interpretations. The collective interpretation of (28b) is not obviously available, and it is often reported in the literature that *every* is a distributive universal or that *every NP* is unacceptable with collective predicates like *gathered* (e.g., Beghelli and Stowell, 1997; Champollion, 2020; Gil, 1995; Tunstall, 1998; Winter, 2002). One has to look to highly infrequent examples to see clear non-distributive readings of *every*, like the differences between the *each* and *every* variants of (29).

(29) Determine whether {each/every} dragon is dangerous.

In particular, the *each* variant of (29) can be read as a request for a pair-list response (e.g., dragon 1: yes, dragon 2: yes, dragon 3: no). But the *every* variant seems to require a single affirmative or negative answer (Beghelli, 1997; Szabolcsi, 2015; Williams, 1986). Still, *every* generally patterns with *each* in terms of being used distributively. But the result from our first two experiments is that *every* and *all* pattern together to the exclusion of *each*.

One could maintain that the root cause of obligatory distributivity for *each* is also what explains our results. For example, Beghelli and Stowell (1997) stipulate that *each* has a strong [+DISTRIBUTIVE] feature not present on *every* and *all*. One way or another, this feature mandatorily gives rise to distributive interpretations (e.g., by forcing the *each NP* to associate with an unpronounced distributive operator that ensures the predicate applies to individuals). This same process might also underlie our findings. In that case, our results would not speak to the representational format of *each* but to the representational format of the distributive feature carried by *each* or of the distributive operator that *each NP* always associates with to ensure that the predicate applies to each individual, not the collection as a whole. In fact, our tentative way of formalizing *each* as restricted first-order quantification is the semantics Szabolcsi (2010) gives for the distributive operator responsible for enforcing distributivity in sentences without an overt *each* (for a treatment of *each* as a pronunciation of the distributive operator, see LaTerza, 2014).

Importantly, we are not suggesting that distributivity and collectivity reduce to the first-order/second-order distinction. Given that *most* is second-order but can be used distributively, such a proposal would be a non-starter. However, *each* being first-order could explain its mandatory distributivity as a direct consequence of its representation (without the need for positing an additional feature). At the same time, *every* and *all* being second-order would cause no more of a problem regarding the availability of distributive readings than *most* being second-order does.

Of course, there are other proposed specifications of *each* that account for, among other facts, its mandatory distributivity. Tunstall (1998) offers a semantics for *each* that forces subevents to be distinct. Winter (2002) argues that *each* and *all* expect predicates of fundamentally different types (atom vs. set). Champollion (2015) differentiates *each* from *all* by specifying associated presuppositions that provide the granularity over which they quantify (e.g., atoms for *each;* small sets for *all*). These proposals include machinery that serves to distinguish individuals and groups, and as such, may be compatible with our results.

Specific linking hypotheses from the details of these proposals to the observed data would need to be provided and tested. The linking hypothesis invoked throughout this paper—that second-order representations are naturally understood as being about groups whereas first-order representations are naturally understood as being about individuals—is straightforward and supported by representational systems independently known to exist in the mind; namely, ensembles and object-files (Kahneman et al., 1992; Ariely, 2001; Feigenson and Carey, 2003; Feigenson et al., 2004; Halberda et al., 2006; Whitney and Yamanashi Leib, 2018; a.o.).

7.2 Could this be about usage, not meaning?

It could be that some fact about these quantifiers' *usage*—as opposed to their meaning—leads to the observed behavioral differences. One challenge in giving usage-based explanations is that there is no clear theory of what kinds of usage facts should matter in what kinds of situations. To take one example, Solt (2016) reports (based on a corpus analysis) that English *most* is very rarely used when the percentage referenced is below 60%. Indeed, this apparent fact can have pragmatic effects on the felicity of *most*-claims in everyday discourse. By all accounts, it seems to be a robust and important detail that people know about *most*.

But despite this knowledge, participants in experimental settings have no problem accepting displays in which 55% of dots are blue as perfectly fine instances of *most of the dots are blue* (see e.g., Pietroski et al., 2009). For any usage-based explanation then, a reasonable linking hypothesis is required that explains why, in a given experiment, the detail in question should matter to the exclusion of other details that evidently do not.

Perhaps one potential usage-centered explanation involves frequency. Uses of *each* are likely less frequent than the other universals. This suggests a plausible linking hypothesis: processing *each* in some way takes extra effort or cognitive resources due to its low frequency, effort which could have otherwise been spent on encoding the cardinality of the restrictor set. Follow-up experiments aimed at ruling out this alternative are underway. Namely, a first-order *each* predicts that tasks probing properties of individual dots instead of ensemble summary statistics should lead participants to perform better following *each*-sentences than *every*- or *all*-sentences. Superior performance following *each*-sentences could not as obviously be explained by excess processing costs incurred for *each*'s low frequency.

7.3 Could this be about the pragmatics of task-switching?

There is a potential concern about the pragmatics of our task: participants may have adopted a different strategy in the second block of the experiment for the sake of doing something new upon being given a new expression to evaluate. It is important to guard against such meaning-independent strategizing. In the future, we intend to run a large-n between-subjects version to alleviate this concern.

For now though, it is worth noting that just saying that participants switch strategies for the sake of change is not enough, on its own, to explain the above results. It is not so much that the quantifier meanings seem to be differentiated by the experiments, but rather the particular direction of their differentiation and its consistency across observers that wants for explanation. If the only constraint at play is the pragmatic one that "one should use different strategies in different situations," then we might expect participants to start out with an individual-based strategy in the first block and shift to a group-based one in the second block or vice versa. Either pattern of performance would result in no difference between the two quantifiers tested in each experiment (since the block order was counterbalanced across participants).

7.4 Might some meanings be representationally neutral?

Throughout this paper, we argued that *each* has a first-order specification whereas *every* and *all* have second-order specifications. We contrasted this with the possibility that all three universals are representationally neutral. On such a view, when speakers acquire a universal quantifier, what they acquire is its truth-conditions, not represented in any particular format (the format, then, might be expected to differ across individuals). Or perhaps they acquire an equivalence class of logically equivalent representations (the format, then, might be expected to differ across contexts).

This does not exhaust the space of logical possibilities, though. Our results could also be accounted for by *each* being first-order (thus driving participants to represent individuals), and *every* and *all* being representationally neutral (thus not biasing any particular verification strategy). Similarly, our data are consistent with *every* and *all* being second-order (thus driving participants to represent groups) and *each* being representationally neutral.

In future work, we hope to distinguish these hypotheses empirically. But at least initially, we're suspicious of the idea that some quantifier meanings are representationally neutral, while others exhibit particular representational formats (shared across speakers). This seems like positing an ad hoc complication, just to preserve as much format-neutrality as the currently available evidence permits. What would meanings be such that some of them are extensions, specified via classes of truth-conditionally equivalent logical forms, while others are more like certain specifications of extensions? As one reviewer put it: if things have representations (which they do), and if these representations matter (which the paper presents evidence for), then there should be no such thing as representational neutrality. We wholly agree.

8 Conclusion

One tradition in semantics treats expressions as names for things in the world: *groundhog* is the name for a set or a function that returns a set given a world and *every* is the name for a mind-independent relation between two sets or a function of type <<e, t>, <<e, t>, t>> (Davidson, 1967; Lewis, 1975; Montague, 1973; a.o.). On this view, meanings are representationally neutral and the formalisms deployed by theorists aren't meant to be related to whatever mental vocabulary humans use to represent the semantic properties of linguistic expressions (Dowty, 1979; see Williams, 2015 for discussion). And for some purposes—e.g., exploring the compositional properties of meanings—it makes sense to abstract away from details about representational formats.

But while it's true that logically equivalent descriptions of meanings are empirically equivalent for purposes of capturing truth conditions, it does not follow that meanings themselves are representationally neutral in the minds of speakers. Here we've argued that despite their truth-conditional equivalence, a specification like (1) or (2) better describes the mental representation of a sentence like *each circle is green* than a specification like (3) or (4) (and that the opposite is true for sentences with *all* or *every*).

(1) $\lambda\Psi \cdot \lambda\Phi \cdot \forall x : \Psi(x)[\Phi(x)]$
(2) $\lambda\Psi \cdot \lambda\Phi \cdot \sim \exists x : \Psi(x)[\sim \Phi(x)]$
(3) $\lambda\Psi \cdot \lambda\Phi \cdot \{x : \Psi(x)\} \subseteq \{x : \Phi(x)\}$
(4) $\lambda\Psi \cdot \lambda\Phi \cdot \{x : \Psi(x)\} = \{x : \Psi(x)\ \&\ \Phi(x)\}$

More generally, the proposed representational difference—between first-order and second-order specifications—embraces the interface between linguistics and psychology in that this distinction grounds out in functions that are available elsewhere in cognition (i.e., systems for representing groups and individuals). We have focused on linguistic verification tasks, which provide one important lens on representational formats. Another lens can be provided by asking whether distinctions between first-order and second-order representations are available to prelinguistic infants. We expect that they will be, and studies to test these predictions are currently underway in our labs.

Even if the concepts are present early in life, we are left with a language acquisition question for future research: how do learners come to associate a first-order meaning with one pronunciation and a truth-conditionally equivalent second-order meaning with another? While the results we've reported here tell us something about how meanings are represented, they urge investigation into what facts about usage would allow different words to be associated with different kinds of representations (for initial suggestions, see Knowlton and Lidz, 2021).

Acknowledgements This work was supported by the National Science Foundation (Grants #NRT-1449815 to T.K. and #BCS-2017525 to T.K. and J.L.) and by the James S. McDonnell Foundation (Grant on *The Nature and Origins of the Human Capacity for Abstract Combinatorial Thought* to P.P., J.H., and J.L.). We are grateful to two anonymous *Linguistics and Philosophy* reviewers for valuable feedback on previous drafts. For helpful discussion throughout this project, we thank: Alexander Williams, Zoe Ovans, Ellen Lau, Darko Odic, Nicolò Cesana-Arlotti, and Adam Liter, as well as audiences at MACSIM 2017, LSA 2019, and CUNY 2019.

References

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest: Akademiai Kiado.

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences, 15*(3), 122–131.

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science, 19*(4), 392–398.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*(2), 157–162.

Baker, M. C. (1995). On the absence of certain quantifiers in Mohawk. In E. Bach, E. Jelinek, A. Kratzer, & B. H. Partee (Eds.), *Quantification in natural languages* (pp. 21–58). Dordrecht: Kluwer.

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

Beghelli, F. (1997). The syntax of distributivity and pair-list readings. In A. Szabolcsi (Ed.), *Ways of scope taking* (pp. 349–409). Dordrecht: Kluwer.

Beghelli, F., & Stowell, T. (1997). Distributivity and negation: The syntax of *each* and *every*. In A. Szabolcsi (Ed.), *Ways of scope taking* (pp. 71–107). Dordrecht: Kluwer.

Boolos, G. (1984). To be is to be a value of a variable (or to be some values of some variables). *The Journal of Philosophy, 81*(8), 430–449.

Burr, D., & Ross, J. (2008). A visual sense of number. *Current Biology, 18*(6), 425–428.

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

Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.

Champollion, L. (2015). Stratified reference: the common core of distributivity, aspect, and measurement. *Theoretical Linguistics, 41*(3–4), 109–149.

Champollion, L. (2020). Distributivity, collectivity, and cumulativity. In D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann, & T.E. Zimmerman (Eds.), *The Wiley Blackwell companion to semantics*. https://doi.org/10.1002/9781118788516.sem021.
Chen, L. (1982). Topological structure in visual perception. *Science, 218*(4573), 699–700.
Chen, L. (2005). The topological approach to perceptual organization. *Visual Cognition, 12*(4), 553–637.
Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
Chomsky, N. (1977). *Essays on form and interpretation*. Amsterdam: North Holland.
Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. New York: Praeger.
Chomsky, N. (1995). Language and nature. *Mind, 104*, 1–61.
Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.
Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43*(4), 393–404.
Church, A. (1941). *The calculi of lambda conversion*. Princeton: Princeton University Press.
Davidson, D. (1967). Truth and meaning. *Philosophy, language, and artificial intelligence* (pp. 93–111). Dordrecht: Springer.
Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.
Dowty, D. (1979). *Word meaning and Montague Grammar*. Dordrecht: Reidel.
Dowty, D. (1987). Collective predicates, distributive predicates, and *all*. In A. Miller & Z. Zhang (Eds.), *Eastern States Conference on Linguistics (ESCOL) 3* (pp. 97–115). Columbus: Ohio State University.
Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Devotional Sciences, 6*, 568–584.
Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307–314.
Feiman, R., & Snedeker, J. (2016). The logic in language: How all quantifiers are alike, but each quantifier is different. *Cognitive Psychology, 87*, 29–52.
Frege, G. (1893). *Grundgesetze der Arithmetik* (Vol. 1). Jena: H. Pohle.
Gil, D. (1992). Scopal quantifiers: Some universals of lexical effability. *Meaning and Grammar: Cross-Linguistic Perspectives, 10*, 303.
Gil, D. (1995). Universal quantifiers and distributivity. In E. Bach, E. Jelinek, A. Kratzer & B. H. Partee (Eds.), *Quantification in natural languages* (pp. 321–362). Dordrecht: Kluwer.
Goodman, N. (1949). On likeness of meaning. *Analysis, 10*(1), 1–7.
Goodman, N. (1953). On some differences about meaning. *Analysis, 13*(4), 90–96.
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., & Feigenson, L. (2008). Developmental change in the acuity of the" number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology, 44*(5), 1457–1465.
Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive internet-based sample. *Proceedings of the National Academy of Sciences, 109*(28), 11116–11120.
Halberda, J., & Odic, D. (2015). The precision and internal confidence of our approximate number thoughts. In D. Geary, D. Berch, K. Mann Koepke (Eds.), *Evolutionary origins and early development of number processing* (pp. 305–333). Amsterdam: Elsevier.
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 development of "most" comprehension and its potential dependence on counting ability in preschoolers. *Language Learning and Development, 4*(2), 99–121.
Halle, M. (1978). Knowledge unlearned and untaught: What speakers know about the sounds of their language. In M. Halle, J. Bresnan & G. A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 294–303). Cambridge, MA: MIT Press.
Higginbotham, J. (1985). On semantics. *Linguistic Inquiry, 16*, 547–593.
Ioup, G. (1975). Some universals for quantifier scope. *Syntax and Semantics, 4*, 37–58.
Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition, 106*, 1221–1247.
Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24*(2), 175–219.

Kazanina, N., Phillips, C., & Idsardi, W. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences, 103*(30), 11381–11386.

Knowlton, T., Hunter, T., Odic, D., Wellwood, A., Halberda, J., Pietroski, P., & Lidz, J. (2021). Linguistic meanings as cognitive instructions. *Annals of the New York Academy of Sciences*. https://doi.org/10.1111/nyas.14618.

Knowlton, T., & Lidz, J. (2021). Genericity signals the difference between *each* and *every* in child-directed speech. In D. Dionne & L. Vidal Covas (Eds.), *Proceedings of the 45th annual Boston University Conference on Language Development (BUCLD)* (pp. 399–412). Somerville: Cascadilla Press.

Knowlton, T., Pietroski P., Williams, A., Halberda, J., & Lidz, J. (2020). Determiners are "conservative" because their meanings are not relations: Evidence from verification. *Proceedings of SALT 30*, 206–226.

Kripke, S. (1980). *Naming and necessity*. Cambridge, MA: Harvard University Press.

Križ, M. (2015). Aspects of homogeneity in the semantics of natural language. Doctoral dissertation, Universität Wien.

Krueger, L. E. (1984). Perceived numerosity: A comparison of magnitude production, magnitude estimation, and discrimination judgments. *Perception and Psychophysics, 35*, 536–542.

Kurtzman, H. S., & MacDonald, M. C. (1993). Resolution of quantifier scope ambiguities. *Cognition, 48* (3), 243–279.

Laming, D. R. J. (1997). *The measurement of sensation*. Oxford: Oxford University Press.

LaTerza, C. (2014). Distributivity and plural anaphora. Ph.D. dissertation, University of Maryland.

Lewis, D. (1975). Languages and language. In K. Gunderson (Ed.), *Minnesota Studies in the Philosophy of Science*, vol. 7 (pp. 3–35). Minneapolis: University of Minnesota Press.

Lewis, D. (1986). *On the plurality of worlds*. Oxford: Blackwell.

Libertus, M. E., Odic, D., & Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychologica, 141*(3), 373–379.

Lidz, J., & Musolino, J. (2002). Children's command of quantification. *Cognition, 84*(2), 113–154.

Lidz, J., Pietroski, P., Halberda, J., & Hunter, T. (2011). Interface transparency and the psychosemantics of most. *Natural Language Semantics, 19*(3), 227–256.

Lu, Z.-L., & Dosher, B. (2014). *Visual psychophysics: From laboratory to theory*. Cambridge, MA: MIT Press.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. MIT Press.

Montague, R. (1973). The proper treatment of quantification in ordinary English. In K. J. J. Hintikka, J. M. E. Moravcsik, & P. Suppes (Eds.), *Approaches to natural language* (pp. 221–242). Dordrecht: Springer.

Odic, D., Im, H. Y., Eisinger, R., Ly, R., & Halberda, J. (2016). Psimle: A maximum-likelihood estimation approach to estimating psychophysical scaling and variability more reliably, efficiently, and flexibly. *Behavior Research Methods, 48*(2), 445–462.

Odic, D., Pietroski, P., Hunter, T., Halberda, J., & Lidz, J. (2018). Individuals and non-individuals in cognition and semantics: The mass/count distinction and quantity representation. *Glossa: a Journal of General Linguistics*, *3*(1). https://doi.org/10.5334/gjgl.409.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., et al. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 1-9.

Phillips, C. (2006). The real-time status of island phenomena. *Language*, 795-823.

Pietroski, P. M. (2018). *Conjoining meanings: Semantics without truth values*. Oxford: Oxford University Press.

Pietroski, P., Lidz, J., Hunter, T., & Halberda, J. (2009). The meaning of 'most': Semantics, numerosity and psychology. *Mind and Language, 24*(5), 554–585.

Pietroski, P., Lidz, J., Hunter, T., Odic, D., & Halberda, J. (2011). Seeing what you mean, mostly. *Experiments at the Interfaces, 37*, 181.

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. https://www.R-project.org/

Rescher, N. (1962). Plurality quantification. *Journal of Symbolic Logic, 27*, 373–374.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464.

Solt, S. (2016). On measurement and quantification: The case of *most* and *more than half*. *Language, 92* (1), 65–100.

Stevens, S. S. (1964). Concerning the psychophysical power law. *Quarterly Journal of Experimental Psychology, 16*(4), 383–385.

Stone, M. (1979). Comments on model selection criteria of Akaike and Schwarz. *Journal of the Royal Statistical Society, Series B, 41*, 276–278.

Szabolcsi, A. (2010). *Quantification*. Cambridge: Cambridge University Press.

Szabolcsi, A. (2015). Varieties of quantification. In N. Riemer (Ed.), *The Routledge handbook of semantics* (chap. 18). London: Routledge.

Tomaszewicz, B. M. (2011). Verification strategies for two majority quantifiers in Polish. *Proceedings of Sinn Und Bedeutung, 15,* 597–612. https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/402.

Travis, C. (2008). *Occasion sensitivity*. Oxford: Oxford University Press.

Tunstall, S. (1998). The interpretation of quantifiers: Semantics and processing. Ph.D. dissertation, University of Massachusetts at Amherst.

Vendler, Z. (1962). *Each* and *every*, *any* and *all*. *Mind, 71*(282), 145–160.

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology, 69*, 105–129.

Wiggins, D. (1980). 'Most' and 'All': some comments on a familiar programme, and on the logical form of quantified sentences. In M. Platts (Ed.), *Reference, truth and reality: Essays on the philosophy of language* (pp. 318–346). London: Routledge & Kegan Paul.

Williams, A. (2015). *Arguments in syntax and semantics*. Cambridge: Cambridge University Press.

Williams, E. (1986). A reassignment of the functions of LF. *Linguistic Inquiry, 17*(2), 265–299.

Winter, Y. (2002). Atoms and sets: A characterization of semantic number. *Linguistic Inquiry, 33*(3), 493–505.