

TITLE: The Mental Representation of Universal Quantifiers

Tyler Knowlton*
Department of Linguistics
University of Maryland, College Park

Paul Pietroski
Department of Philosophy
Rutgers, The State University of New Jersey-New Brunswick

Justin Halberda
Department of Psychological and Brain Sciences
Johns Hopkins University

Jeffrey Lidz
Department of Linguistics
University of Maryland, College Park

*Correspondence should be addressed to

Tyler Knowlton
1401 Marie Mount Hall
College Park, MD 20742
tzknowlt@umd.edu

Keywords: natural language quantifiers; semantics; meaning; logic

ABSTRACT

A sentence like *every circle is blue* might be understood in terms of individuals and their properties (e.g., for each thing that's a circle, it's blue) or in terms of a relation between groups (e.g., the blue things include the circles). Formally, theorists can use the tools of first-order logic and/or the more powerful tools of second-order logic to specify the contents of universally quantified statements. We offer new evidence that this formal distinction is psychologically realized and has behavioral repercussions. Participants were shown displays of dots and asked to evaluate statements 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 *every*- or *all*-statements compared to *each*-statements. This suggests that *every*- and *all*-statements are mentally represented in second-order terms that highlight comparison of groups whereas *each*-statements are represented in first-order terms that focus on individuals. Since the statements that participants evaluate are truth-conditionally equivalent, our results also bear on questions concerning how meanings are related to truth-conditions.

1. Introduction

Words connect pronunciations with meanings that support communication. For example, speakers can have conversations about rabbits despite variation in how they think about rabbits. This suggests that the word *rabbit* has a sharable meaning that is somehow linked to the relevant animals. Semanticists often capture this idea by describing meanings as abstract entities like the set of (all and only) the rabbits or the function that maps each possible world **w** to the set of rabbits at **w**. On this view, speakers can represent rabbits in different ways, yet still connect the pronunciation of *rabbit* to a common meaning.

The meaning of a quantifier like *every* might also be an abstract entity: a relation between two groups (e.g., the rabbits and the furry things). This relation can be described equally well in many different ways. Consider the options in (1)-(4), where R and F stand for predicates like *rabbit* and *furry*.

- | | |
|---------------------------------|---|
| (1) $\forall x:Rx(Fx)$ | \approx for each thing that's a rabbit, it's furry |
| (2) $\sim\exists x:Rx(\sim Fx)$ | \approx it's not the case that some rabbit isn't furry |
| (3) $R \subseteq F$ | \approx the set of rabbits is a subset of or is identical to
the set of furry things |
| (4) $R = R \cap F$ | \approx the set of rabbits is identical to the set of furry rabbits |

These options are logically equivalent. Each condition in (1)-(4) is met just in case the other three are. If the meaning of *every* is the format-neutral relation that can be described in any of these ways, competent speakers might differ in how they represent this meaning, and in how they represent the meanings of sentences like *Every rabbit is furry*. Conversely, all speakers might mentally encode the relevant relation in one particular way. If so, this would suggest that the meaning of *every* is more fine-grained and is better thought of as a particular instantiation of the relation (e.g., (3)). On this view, the meaning of *every* would be a relation

specified in a particular format, as opposed to the relation itself (see e.g., Frege, 1893; Church, 1941; Tichý, 1969; Marr 1982; Horty 2007; Pietroski, 2018; also see Geurts & Nouwen, 2007 for an argument that *at least three* and *more than two* have different specifications despite them expressing the same relation).

Here we argue that the meanings of *each*, *every*, and *all* are not neutral with respect to representational format. In particular, we focus on the distinction between first- and second-order specifications (described in section 1.1) and argue that this formal distinction is psychologically realized in a way that has behavioral repercussions.

Our main prediction is that if speakers understand a quantifier *Q* in a second-order way, then the phrase *Q big dot(s)* should facilitate representing the big dots in a way that supports encoding collective properties like cardinality (no individual dot is *four*). 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 statement like *Q big dots are blue*, relative to a perceived scene, and then asking a follow-up question like *how many big dots were there?*

In order to demonstrate the distinct behavioral patterns that occur in response to group-comparison and individual-based meanings, we first show that provably second-order *most*-statements promote better knowledge of cardinality compared to simple existential-statements (which do not require a second-order treatment). Then we report a similar pattern for statements with *all* and *every* in comparison to statements with *each*. We argue that this provides evidence that *all* and *every* are represented in a second-order format, whereas *each* is represented as first-order. To be sure, these results might also reflect whichever other semantic differences underlie the well-known fact that *each* – unlike the other universals – mandatorily gives rise to distributive interpretations (e.g., Vendler, 1962; Dowty, 1987). We address this possible explanation in section 5.

1.1 Classifying individuals versus comparing groups: a primer on first- and second-order logic

The specifications in (1) and (2) are first-order whereas the specifications in (3) and (4) are second-order. Informally, the difference between first- and second-order specifications is this: first-order specifications characterize the relation in terms of individual rabbits and whether they're furry; second-order specifications do so by comparing the rabbits (taken together) with the furry things (taken together).

Formally, first-order logic makes use of logical constants, like “*r*” in (5) and (6), and properties that can be predicated of those individuals, like “Rabbit”.

(5) Rabbit(*r*) ≈ *r* is a rabbit

(6) $r \in \{x: \text{Rabbit}(x)\}$ ≈ *r* is a member of the set of rabbits

Additionally, first-order specifications can relate two logical constants, e.g., with the “Chased” relation in (7). First-order specifications can also employ quantificational devices that range over logical constants, like the “ $\forall x$ ” in (8).

(7) Chased(*r*, *y*) ≈ *r* chased *y*

(8) $\forall x: \text{Rabbit}(x)[\text{Chased}(x, y)]$ \approx for each thing that's a rabbit,
it chased y

But first-order specifications can't relate positions occupiable by predicates like "Rabbit" or "Chased". In other words, first-order resources are limited in ways that preclude quantification into, or otherwise binding, predicate positions.¹ By contrast, second-order specifications allow quantification into predicate positions, as seen in (9) and (10), where "R" might stand for "Rabbit" or any other property and "C" might stand for "Chased" or any other two-place relation.

(9) $\exists R: R(\text{peter})$ \approx there's some property, R, that applies to peter
(10) $\exists C: C(\text{peter}, \text{ben})$ \approx there's some relation, C, that holds between
peter and ben

Expressing certain contents requires the tools of second-order logic. Consider the second-order (11), which indicates that the rabbits (R) and the dogs (D) correspond one-to-one.

(11) $\exists R \exists D \{ \text{OneToOne}(R, D) \ \& \ \forall x [Rx \equiv \text{Rabbit}(x)] \ \& \ \forall x [Dx \equiv \text{Dog}(x)] \}$

This specification allows for comparing the rabbits (taken together) with the dogs (taken together). Such a comparison cannot be made with only first-order resources (i.e., only by relating individuals that are rabbits to individuals that are dogs).

Turning to natural language, the proportional quantifier *most* – as in *most rabbits are furry* – has a meaning that cannot be specified in first-order terms (Rescher, 1962; Wiggins, 1980; Barwise & Cooper, 1981). This is because *most rabbits are furry* is, one way or another, a claim about the ratio of rabbits to furry rabbits, which cannot be reduced to claims about individuals and their properties. The *most* relation can be specified in various second-order ways though, including (12)-(14).²

(12) $|R \cap F| > |R \cap \text{Complement}(F)|$

¹ A specification like (1*) that treats the predicates "R" and "F" as sets is nonetheless first-order, since those predicate positions are not being related to each other. Likewise, a specification like (3*), where R and F do not correspond to sets and no set-theoretic notion is deployed, is still second-order.

(1*) $\forall x: x \in R(x \in F)$ \approx each element of R is an element of F
(3*) Among(the Rs, the Fs) \approx the Rs are among the Fs

In general, appeal to sets – or other "plural entities" like mereological sums or lattices (Scha, 1984; Link, 1983; Schwarzschild, 1992, 1996; Chierchia, 1998) – can be replaced with plural quantification (though subtle distinctions arise when considering vagueness, predicates like *set that is not an element of itself* and differences in the logical implications of sentences like *There are some rocks in the box* and *There is a set of rocks, and it is in the box*; see Boolos, 1984, 1999; Schein, 1993; Pietroski, 2005, 2018).

² For empirical discussion of (12)-(14), see Halberda, Taing, & Lidz (2008), Hackl (2009), Pietroski et al. (2009), and Lidz et al. (2011), who offer evidence in favor of a specification like (13).

$$(13) |R \cap F| > |R| - |R \cap F|$$

$$(14) |R \cap F| > \frac{1}{2}(|R|)$$

In this respect, *most* is similar to terms like *ninety percent*, which also depend on ratio. Namely, nine out of ten counts as *ninety percent* but nine out of fifty does not. Because the quantity of the sets being compared matters, *ninety percent* is not first-orderizable.

But the relation of universal quantification isn't proportional, and it can be described in first-order terms. So it isn't obvious that *each*, *every*, and *all* are like *most* and *ninety percent* in having meanings that speakers somehow specify in second-order terms. In this respect, the universals are similar to numeric relations like *four*, which is first-orderizable. The second-order (15) is logically equivalent to the first-order but lengthy (16).

$$(15) |R \cap F| = 4$$

$$(16) \exists x \exists y \exists z \exists w \{x \in R \ \& \ y \in R \ \& \ z \in R \ \& \ w \in R \ \& \ x \in F \ \& \ y \in F \ \& \ z \in F \ \& \ w \in F \ \& \\ (x \neq y) \ \& \ (x \neq z) \ \& \ (x \neq w) \ \& \ (y \neq z) \ \& \ (y \neq w) \ \& \ (z \neq w) \ \& \\ \forall s: s \in R \ \& \ s \in F [(x = s) \vee (y = s) \vee (z = s) \vee (w = s)]\}$$

It might seem implausible that *four* is represented as the cumbersome (16) in the minds of speakers, but it is an empirical question whether the meaning of *four* is more like (15), more like (16), or more like the representationally neutral shared content of them both. Likewise, for the universal quantifiers, the relevant relation can be captured with either type of description. And it is an empirical question whether the meaning of *every* is more like (1) and (2), more like (3) and (4), or more like the representationally neutral shared content of (1)-(4). In other words, is the distinction between first- and second-order specifications purely formal or does it correspond to a genuine psychological distinction in the minds of speakers?

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

Before turning to the experiments, it is 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-/second-order distinction, though without yet providing any clear evidence of a corresponding psychological distinction. So it may come as no surprise that the distinct words for expressing universal quantification have distinct meanings. But 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 meaning in different ways.

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

For example, even if (17) and (18) are truth-conditionally equivalent, (17) seems to be about classifying one kind of thing (rabbits) as falling under a broader category of things

(animals), whereas (18) invites the thought that the property of being an animal applies to each thing we're willing to call a rabbit.

(17) All rabbits are animals.

(18) Each rabbit is an animal.

Likewise, although (19) is not about classifying anything, using *each* 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 & Stowell, 1997).

(19) 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 (20a). On the other hand, *all* is easily used with many collective predicates³; though it doesn't force collective interpretations, as illustrated with the distributive (20b).

(20) a. {*Each / All} of the soldiers surrounded the fortress.

b. {Each / All} of my students can sing that song well as a solo piece.

Grammatically, it seems that *each* requires the partitive *of* when used with a plural noun phrase like *the dogs* (21a), whereas *all* can take the plural noun phrase *the dogs* or the full prepositional phrase *of the dogs*.

(21) a. Each of the dogs barked. / *Each 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 (cf. (22), which seems synonymous with (21b)).

(22) The dogs all barked.

It may be more significant that *one* can be freely added to sentences with *each*, without a change in meaning, as in (23a). But with *one*, (23b) is highly marked, and a special prosody and/or context is required to convey its meaning (namely, that there is only one dog in the domain and it barked).

(23) a. Each (one) of the dogs barked.

³ There is a class of collective predicates that cannot be used with any of the universal quantifiers, like *be numerous*, *be a good team*, and *be a group of five* (Dowty, 1987; Winter, 2002; Champollion, 2015). But given that such predicates run into conflict with multiple quantifiers (including the admittedly second-order *most*) this point can't be used to differentiate among first- and second-order quantification.

- 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 (24); see Vendler (1962) and Dowty (1987).

- (24) a. {*Each/?Every} dot is alike.⁴
b. All dots are alike.

Likewise, *every* requires a singular count noun, as in (25a). But it differs from *each* and *all* in being incompatible with the partitive without support by *one*, as seen in (25b).

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

There are, however, a few respects in which *every* patterns with *all*. Beghelli & 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 (26) and (27).

- (26) a. One kid ate almost {all the cookies/every cookie}
b. *One kid ate almost each cookie
(27) 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*, as we saw above, is not. To take another example, even though (28a) seems like a distributive generalization that is TRUE iff (28b) is TRUE, (28c) carries no generic implication of the sort conveyed with (28a) or (28b).

- (28) a. Every rabbit hops.
b. All rabbits hop.
c. Each rabbit hops.

These facts suggest differences in the syntax and semantics of *each*, *every*, and *all*. A theme throughout these examples (and one often discussed in the literature) is that sentences with *each* are sentences about individuals in some sense and that the same is not always true for sentences with *all* or *every*. Here, we offer independent psychological evidence of this same fact. Our proposed distinction between first- and second-order representations is one way to account for this particular difference. No two-way distinction has any hope of explaining all of

⁴ 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 section 5.

the differences between the universal quantifiers, and the one offered here is far from a complete account of their meanings. But given its independent psychological plausibility and the evidence presented below, we think this distinction between formats is a good place to start.

2. Motivating a Psychophysical Investigation of Quantifier Meanings

One source of evidence for a format distinction can come from looking at how participants verify different quantificational statements. Lidz et al. (2011) argue that all else equal, people are biased to evaluate a given statement with a procedure that transparently reflects that statement's meaning.⁵ This linking hypothesis is in the spirit of Marr (1982): the format of a given thought highlights certain information, making some ways of evaluating that thought more natural than others (see Pietroski et al., 2011).

For example, Pietroski et al. (2009) show that participants verify *most*-statements with a cardinality-based strategy even when superior alternatives are readily available. Namely, participants show clear signatures of using the Approximate Number System (for review, see Feigenson, Dehaene, & Spelke, 2004; Dehaene, 2011) when asked whether *most of the dots are blue*. But for Pietroski et al.'s displays, this was a sub-optimal strategy. When given identical displays and asked to *find the leftover dot*, participants used a one-to-one correspondence strategy that is faster and more accurate than relying on cardinality comparisons. The intuition here is that, if nothing else can explain the change in strategy, the change in meaning must be to blame. Specifically, the representational format: *most* is specified in terms of cardinality, not correspondence. Cross-linguistic work confirms that the predictions of this hypothesis are borne out for majority quantifiers in Polish (Tomaszewicz, 2011) and Cantonese (Knowlton et al., in prep).

Our current question is not about cardinality versus correspondence, but about another format distinction: first- versus second-order representations. As discussed in section 1.1, a first-order format highlights individuals, whereas a second-order format privileges groups. This format distinction should also contribute to determining what verification strategy is used. If a given quantifier has a first-order format, then evaluating statements with it should, all else equal, direct attention to individuals and cause participants to represent individuals as such. On the other hand, if a given quantifier has a second-order format, then evaluating statements with that quantifier should, all else equal, direct attention to groups and cause participants to represent groups as such.

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 & Treisman, 2003; Halberda, Sires, & Feigenson, 2006; Burr & Ross, 2008; Alvarez, 2011; Whitney & Leib, 2018). Representing a group and encoding knowledge of its summary statistics does not even require explicitly representing each

⁵ Their claim is not that meanings *are* verification strategies or that people *always* use a certain strategy to evaluate a given statement. To be sure, there are other considerations that go into determining what verification strategy an individual might deploy in a certain situation. The claim is merely that the format that the meaning is expressed in contributes a detectable influence over verification procedures. And if the 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.

individual constituting that group (Ariely, 2001; Alvarez & Oliva, 2008). On the other hand, merely attending to and representing individuals plausibly will not afford the same access to summary statistics of any groups those individuals belong to, a prediction we test below.

In this paper, we focus on one such statistic: cardinality. The important point for our task is that attending to and representing a group enhances sensitivity to that group's cardinality, compared to attending to and representing the individuals constituting that group. In the four experiments reported here, we use participants' knowledge of cardinality to infer whether they were using a group- or individual-based verification strategy during evaluation. In each trial, participants were shown dot-displays (e.g., Figure 1) and asked to evaluate sentences, like *all of the big dots are blue*.

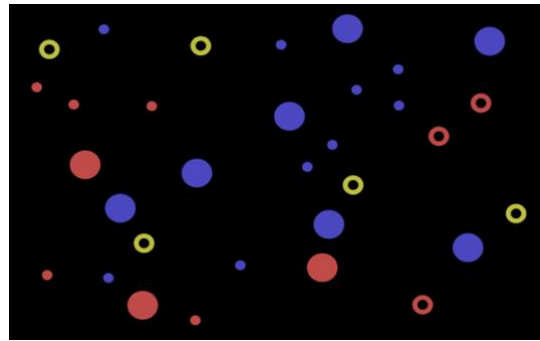


Figure 1: an example of the dot-displays used in the experiments.

After answering “true” or “false” (in this case, the statement is 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?”. We expect participants to have relatively poor estimates in these cases – perhaps answering “7” instead of “14” – since the set of small dots is irrelevant to whether or not *all of the big dots are blue*.

The more important follow-up questions probed the “target” set: “how many big dots were there?”. If *all* has a second-order format, we should 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 the cardinality of that group (in this case, 10). If *all* has a first-order format, we should instead expect them to consider individual dots (and decide whether both predicates – *big* and *blue* – apply). In using this strategy, they should have a worse estimate of the big dots. Our main diagnostic then, will be how well participants know the cardinality of the relevant group after evaluating some quantificational statement. Across four experiments, we use this diagnostic to test the six kinds of statements listed in (29).

- (29) a. Most of the big dots are blue
- b. There is a big dot that's blue
- c. Each of the big dots are blue
- d. Each big dot is blue
- e. Every big dot is blue
- f. All of the big dots are blue

Experiment 1 offers a proof-of-concept by showing that second-order statements like (29a) lead to better knowledge of the relevant set's cardinality (the number of big dots) than plausibly first-order statements like (29b). While *most* is provably second-order, anything that can be represented with first-order logic can be represented with second-order logic as well. This makes it an empirical question whether existential statements like (29b) have first-order formats. Likewise, it is an empirical question whether this format will lead to differential performance on cardinality questions (e.g., it might instead be that any statement in (29) leads participants to represent the set of big dots). And while past results (e.g., Pietroski et al., 2009; Hackl, 2009; Lidz et al., 2011) make it likely that statements with *most* encourage participants to attend to and represent groups, we know of no experiment that has asked participants to explicitly recall cardinality information following a verification task. Experiment 1 aims to answer these three questions.

Using the same methods, experiments 2-4 pit the universal quantifiers *each*, *every*, and *all* against one another. To foreshadow our results, we find that statements with *every* and *all* (like (29e-f)) lead participants to have better estimates of the relevant sets' cardinalities than statements with *each* (like (29c-d)). We find this result despite the fact that the truth-conditions and images used are identical. Given that there is no other reason to switch strategies across the universal quantifiers, we argue these results reflect a difference in how meaning is mentally represented; specifically, the proposed first-/second-order distinction.

3. Experiment 1: *Most of the-* vs. Existential-statements

3.1 Method

3.1.1 Subjects

28 University of Maryland undergraduates participated in exchange for course credit. Three participants were excluded for failing to finish both blocks within the allotted hour and one participant was excluded for performing near chance on the TRUE/FALSE portion. This left us with the desired 24 participants.

3.1.2 Materials

Statements had one of two forms: *most of the [size] dots are [color]* or *there is a [size] dot that's [color]*. Dot displays consisted of red, yellow, and blue dots that could be big, medium, or small (see Figure 1). Medium dots had black holes in the middle, to make them more distinguishable from the other two sizes (Chen, 1982; 2005). Participants were shown all three dot sizes during the training portion of the experiment to ensure that they could be correctly identified.⁶

⁶ Additionally, some participants were given an image of the "target" dot type along with the statement. For example, if the statement was "most of the big dots are blue" they would see a single big blue dot below the text. We thought this might guard against participants misreading the statements and, in general, make the experiment easier. We found no significant difference in performance on accuracy of target trials when these "reminder dots" were present though (*most of the*: $t_{21.3} = 0.87$, $p = .395$; *there is a*: $t_{21.1} = 1.23$, $p = .233$), so we collapsed across the two groups in the analysis presented here and discontinued use of "reminder dots" in all following experiments.

Each display consisted of between 24 and 48 dots, and six size/color combinations were always shown on the screen at once. 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 sets and the superset in parallel (Halberda, Sires, & Feigenson, 2006). Each subset present (e.g., big blue dots) 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. Color and size were randomized across trials. 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).

3.1.3 Procedure

Participants completed 272 trials in a 2 (quantifier) x 2 (set probed) design. In each trial, they read a quantificational statement. Type of quantificational statement was blocked and initial condition was counterbalanced such that 12 participants started with *most of the*-statements and 12 started with *there is a*-statements.

After reading the statement and pressing ‘space’, participants viewed a dot display and evaluated whether that statement 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 judgement. 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’ (Figure 2). All experiments reported here were built using PsychoPy2 (Peirce et al., 2019).

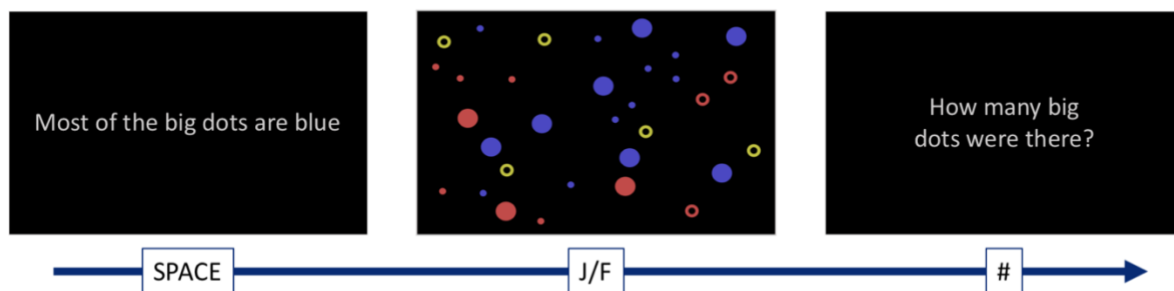


Figure 2: Trial structure of the experiments.

3.2 Predictions

We predict that participants should be better at estimating the cardinality of the “target” set following *most*-statements compared to existential-statements. If a participant is asked to evaluate *most of the big dots are blue* for example, they should have a good estimate of the number of big dots. If they are asked to evaluate *there is a big dot that’s blue*, on the other hand, they should have a worse estimate of the number of big dots. Regardless of the quantifier, we expect participants to have poor guesses of the cardinalities of “distractor” sets: small dots or medium dots, in this example.

We operationalized cardinality knowledge by fitting participants' responses with the standard psychophysical model of cardinality estimation (see the supplementary materials for details). This model includes a parameter for accuracy (β) (e.g., Stevens, 1964) and precision ($1 - \sigma$) (e.g., Laming, 1997). In these terms, we predict that participants will be more accurate and more precise at estimating the cardinality of the “target” set following *most of the*-statements than following *there is a*-statements and that the choice of quantifier will have no such effect on participants' accuracy or precision for “distractor” sets.

3.3 Results & discussion

We fit an accuracy (β) and precision ($1 - \sigma$) to each combination of set probed (“target” and “distractor”) and quantifier (*most of the* and *there is a*). Accuracy and precision were fit simultaneously using the PsiMLE package (Odic et al., 2016) for R (R Core Team, 2017). For this and all subsequent experiments, plots of the raw data and information about performance on the TRUE/FALSE portion are included in the supplementary materials. Here, we focus on comparing the parameter values following *most of the*- and *there is a*-statements. Specifically, we computed differences between the pairs of parameters in Figure 3 and compared each difference to the null result (i.e., 0) of no difference between groups with a Wald test.⁷

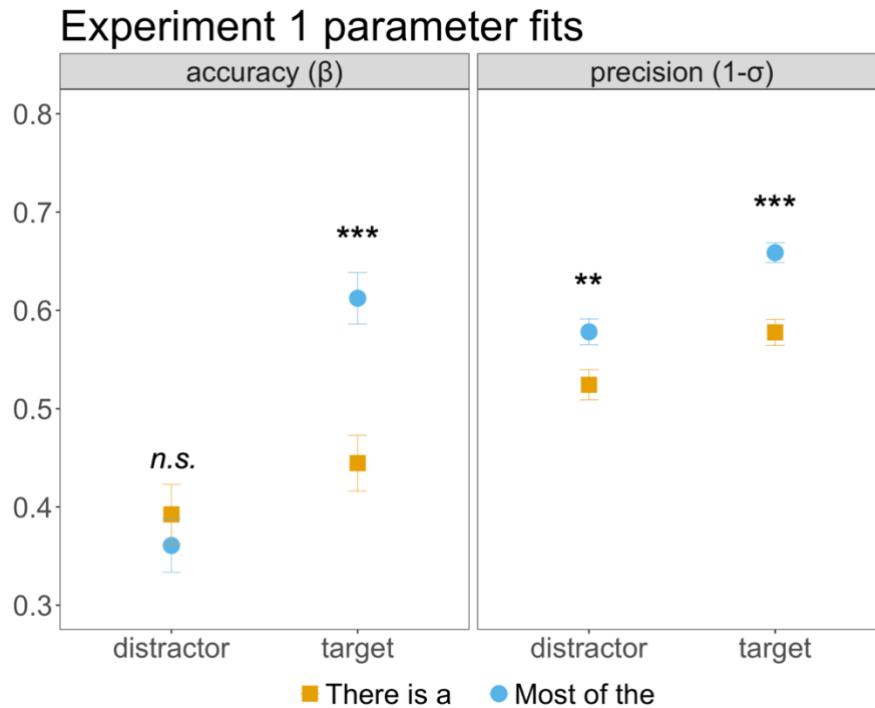


Figure 3: Parameter fits for participants' grouped responses on the cardinality portion of experiment 1. “Target” questions probed the restrictor set of the quantificational statement (e.g., *big dots* in *most of the big dots are blue*); “distractor” questions probed the complement of the restrictor set (e.g., *small or medium dots* in the above example). Precision ($1 - \sigma$) is plotted instead of variability (σ), so that higher values along both dimensions correspond to better estimates of the relevant cardinality.

⁷ Interactions between quantifier and set probed were not tested directly, as we know of no way of computing that statistic for single parameter fits.

Participants were more accurate ($\chi^2 = 18.79$, $p < .001$) and more precise ($\chi^2 = 23.82$, $p < .001$) on target questions following *most of the*-statements than *there is a*-statements. This confirms our predictions that the meaning of *most of the*-statements will lead participants to represent the relevant group whereas the meaning of *there is a*-statements will not. Further, as expected there was no such accuracy boost on distractor questions ($\chi^2 = 0.6$, $p = .439$), but this is qualified slightly by a significant boost in the precision of for cardinality answers on distractor questions following *most of the*-statements ($\chi^2 = 7.14$, $p = .008$). However, without a corresponding boost in accuracy, this increase is hard to interpret. Moreover, in follow-up experiments comparing *most* against existential-statements (see supplementary materials), this effect does not replicate, suggesting it may be spurious.

This result offers a proof-of-concept that some quantificational statements cause participants to represent groups (and to encode group properties like cardinality), whereas others direct attention to individuals (and thus result in degraded knowledge of group properties like cardinality).

One reviewer pointed out an alternative explanation: When evaluating *most of the big dots are blue*, the number of big dots is vitally important as it is directly used in providing an answer. In a statement like *there is a big dot that's blue*, number information is incidental and there is no reason for participants to hold it in memory. The asymmetry observed here may thus be due to a disparity in the usefulness of number information. Follow-up experiments probing other summary statistics (average size and center of mass) are underway, in part to address this concern. If a distinction between first- and second-order representational formats is responsible for these results, as we've proposed, then *most*-statements should lead to superior knowledge of all summary statistics, not just cardinality.

Having established that the format of a quantifier's meaning can impact memory for group summary statistics, we now turn to the first-/second-order distinction in the domain of universal quantifiers.

4. Experiment 2: *Each of the* vs. *All of the*

4.1 Method

4.1.1 Subjects

30 University of Maryland undergraduates participated for course credit. 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 the desired 24 participants.

4.1.2 Materials

Statements had one of two forms: *each of the [size] dots are [color]* or *all of the [size] dots are [color]*. Aside from the statements, the only difference between experiment 2 and experiment 1 was that displays contained fewer "target" dots, on average. Specifically, the average cardinality for the target set was reduced to 8.9 in experiments 2-4, compared to 11.6

in experiment 1. This was done to reduce crowding thus making the cardinality questions easier.⁸

4.1.3 Procedure

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

4.2 Predictions

Given some of the linguistic facts reviewed in section 2, *each* is the most likely of the universals to have a first-order representation (i.e., in terms of categorizing individuals). On the other hand, *all* is the most likely universal to be mentally specified in a second-order way (i.e., in terms of a relation between groups). If this is right, we should expect the following: better accuracy and better precision on target questions following *all*-statements than following *each*-statements but no difference in accuracy or precision on distractor questions.

4.3 Results & discussion

These predictions were borne out (Figure 4): participants were more accurate ($\chi^2 = 7.38$, $p = .007$) and more precise ($\chi^2 = 6.67$, $p = .009$) on target questions following *all*-statements than *each*-statements. As predicted, quantifier had no effect on accuracy ($\chi^2 = 1.36$, $p = .243$) or precision ($\chi^2 = 0.03$, $p = .873$) for questions probing a distractor set.

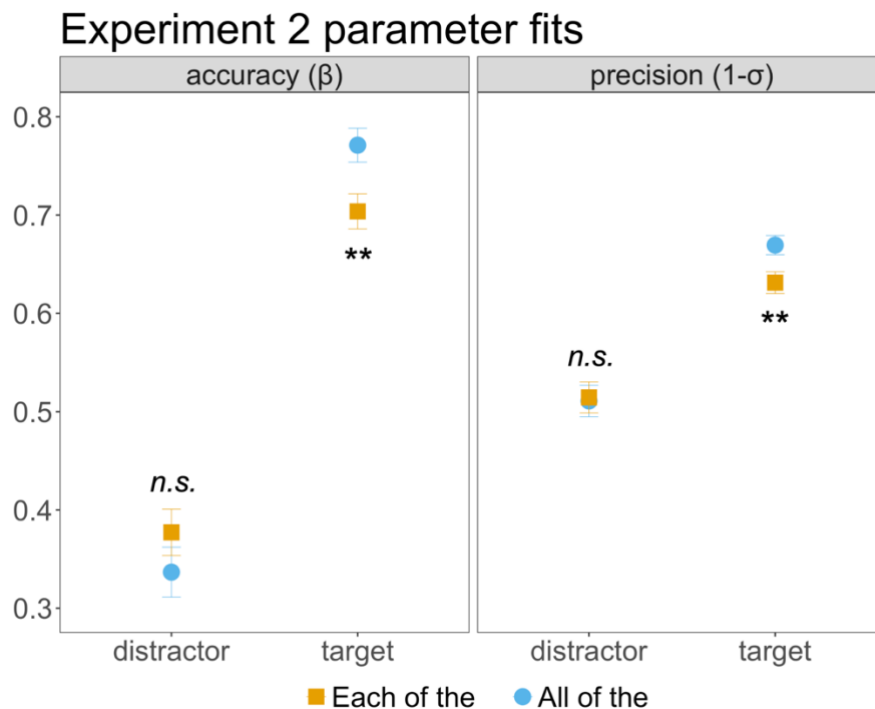


Figure 4: Parameter fits for participants' grouped responses on the cardinality portion of experiment 2. "Target" questions probed the restrictor set of the quantificational statement (e.g., *big dots* in *all of the*

⁸ Estimates for accuracy (β) in experiment 1 were far lower than the "best" possible performance of .8 (e.g., Krueger, 1984; Odic et al., 2016). We suspected that this might be due to the displays being slightly crowded, which Im, Zhong, & Halberda (2016) have shown leads to increased underestimation.

big dots are blue); “distractor” questions probed the complement of the restrictor set (e.g., *small* or *medium dots* in the above example).

That said, participants’ accuracy on target questions following *each*-statements is relatively high compared to the trials in which they were asked about a distractor size. A possible explanation is that participants attended to and represented the individual dots following *each*-statements but nonetheless got some information about the target set during evaluation through other means. For example, they may have adopted a strategy of using viewing time as a proxy for how many relevant dots there were. Or they may have learned to pay attention to the restrictor set as the experiment went on. Since the dots stayed on the screen until participants responded, they might have learned to delay their TRUE/FALSE answer until they were more certain of some sets’ cardinalities.

To test this possibility, we fit new models to participants’ responses from only the first trial of each question type (Figure 5). Again, we found a significant difference between accuracies following target questions ($\chi^2 = 4.93$, $p = .026$) but no difference on distractor trials ($\chi^2 < 0.01$, $p = .995$). Importantly, fitting the model to just the first trial of each type yielded an accuracy for *each* on target trials exactly as bad as we would expect given the results of prior work (Halberda, Sires, & Feigenson, 2006) and a baseline number-estimation experiment using the same stimuli (see supplementary materials).

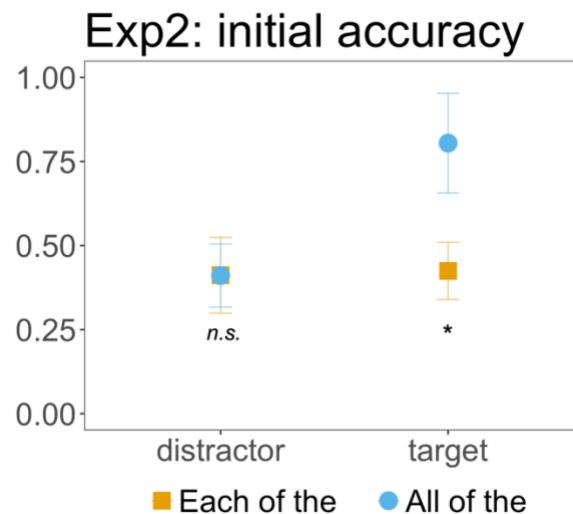


Figure 5: Accuracy (β) fits for each participant’s first trial of each question type.

Having established a difference between *each*-statements and *all*-statements consistent with the hypothesis that *each* is first-order and *all* is second-order, we now turn to *every*. As we saw in section 1, in some linguistic contexts, *every* patterns with *each*, while in other contexts, it patterns with *all*. A task comparing *each* and *every* will allow us to diagnose this aspect of *every*’s meaning.

5. Experiment 3: *Each* vs. *Every*

5.1 Method

5.1.1 Subjects

30 University of Maryland undergraduates participated in exchange for course credit. 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 the desired 24 participants.

5.1.2 Materials

Statements had one of two forms: *each [size] dot is [color]* or *every [size] dot is [color]*. Aside from the statements, the materials were identical to those used in experiment 2.

5.1.3 Procedure

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

5.2 Predictions

If *every* has a second-order representation, we should expect better accuracy and better precision on target questions following *every*-statements than following *each*-statements. If *every* instead has a first-order representation, we should expect choice of quantifier to make no difference to accuracy or precision on target questions. And in any case, we predict that quantifier should make no difference in accuracy or precision on distractor questions.

5.3 Results & discussion

The results match the predictions of *every* being second-order (Figure 6). Accuracy was better on target trials following *every*-statements compared to *each*-statements ($\chi^2 = 25.27$, $p < .001$), as was precision ($\chi^2 = 5.01$, $p = .025$). We observed no boost for accuracy ($\chi^2 = 0.17$, $p = .68$) or precision ($\chi^2 = 0.001$, $p = .97$) on the distractor trials.

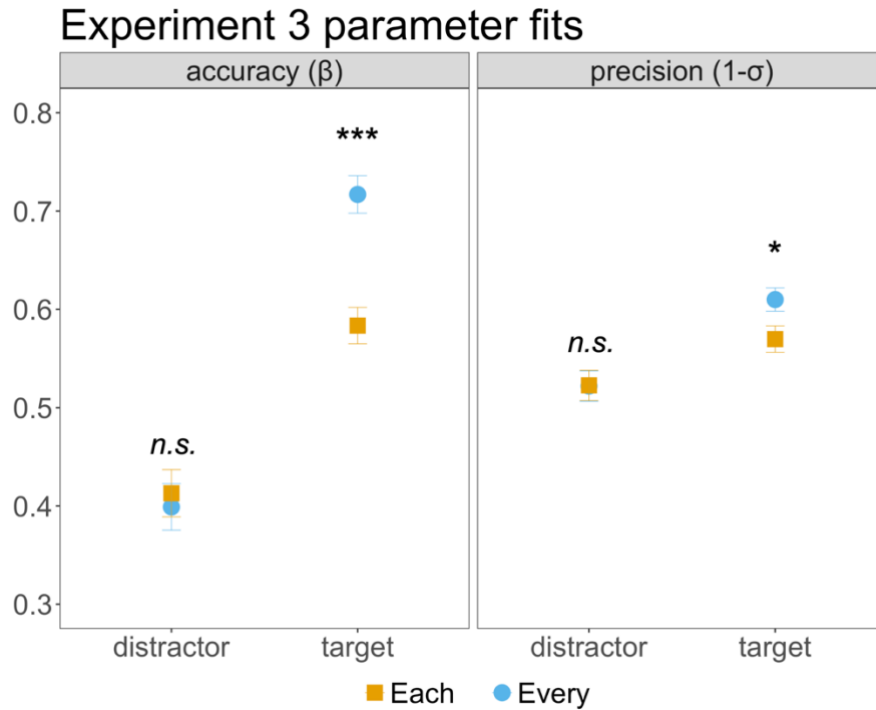


Figure 6: Parameter fits for participants' grouped responses on the cardinality portion of experiment 2. "Target" questions probed the restrictor set of the quantificational statement (e.g., *big dots* in *every big dot is blue*); "distractor" questions probed the complement of the restrictor set (e.g., *small or medium dots* in the above example).

Considering only the first trial of each type yields similar results (Figure 7): accuracy on target trials was significantly better following *every*-statements than following *each*-statements ($\chi^2 = 6.11$, $p = .013$), but there was no difference on distractor trials ($\chi^2 = 0.91$, $p = .339$). As before, we find that accuracy on the first *each* target trial is where we would expect it to be if participants didn't know the cardinality better than they did in the baseline experiment (reported in the supplementary materials).

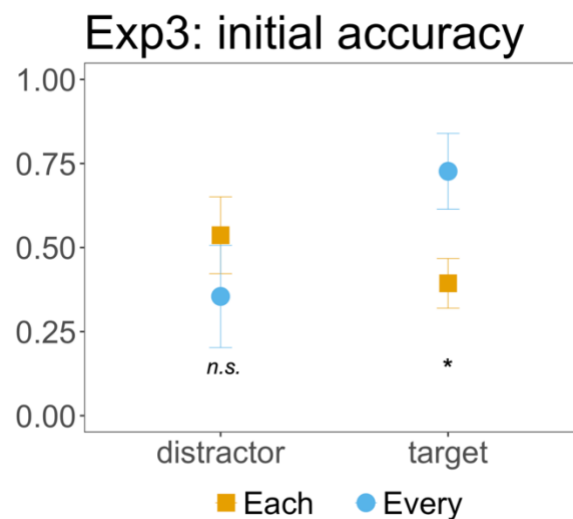


Figure 7: Accuracy (β) fits for each participant's first trial of each question type.

These results support the hypothesis that *every* has a second-order meaning and replicate the earlier finding that *each* has a first-order meaning. To show that this difference is not the result of *every* being contrasted with *each*, experiment 4 compares *every* and *all* directly. If *every* has a second-order meaning, then we should find no difference between *every*- and *all*-statements.

6. Experiment 4: *Every* vs. *All of the*

6.1 Method

6.1.1 Subjects

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

6.1.2 Materials

Statements had one of two forms: *every [size] dot is [color]* or *all of the [size] dots are [color]*. Aside from the statements, the materials were identical to those used in experiments 2 and 3.

6.1.3 Procedure

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

6.2 Predictions

The results of experiments 2 and 3 suggest that *each*-statements rely on first-order representations whereas *all*- and *every*-statements rely on second-order representations. If this is true, then both *every* and *all* should lead to similar performance when compared directly. In particular, we predict no significant differences between accuracy and precision depending on the quantifier used.

6.3 Results & discussion

As seen in Figure 8, we find that both quantifiers lead participants to similar accuracies following target questions ($\chi^2 = 2.95$, $p = .086$). We do find a difference in precision though: participants were more precise at answering target questions following *all*-statements ($\chi^2 = 11.74$, $p < .001$). And while there was no difference in precision on distractor questions ($\chi^2 = 0.25$, $p = .615$), we do find an unexpected difference in accuracy such that participants were more accurate following *every*-statements ($\chi^2 = 25.27$, $p < .001$).

Results from considering just the initial trials are more in line with our predictions: we find no significant accuracy differences between the quantifiers for either target ($\chi^2 = 1.84$, $p = .175$) or distractor ($\chi^2 = 0.32$, $p = .574$) questions. This suggests that both *all* and *every* initially bias participants to attend to and represent groups during evaluation. This is despite the fact that our *every* statements were singular (*every dot*), perhaps discouraging set representations.

Likewise, our *all*-statements were both plural and partitive (*all of the dots*), both of which one might naively expect would encourage attention to groups.

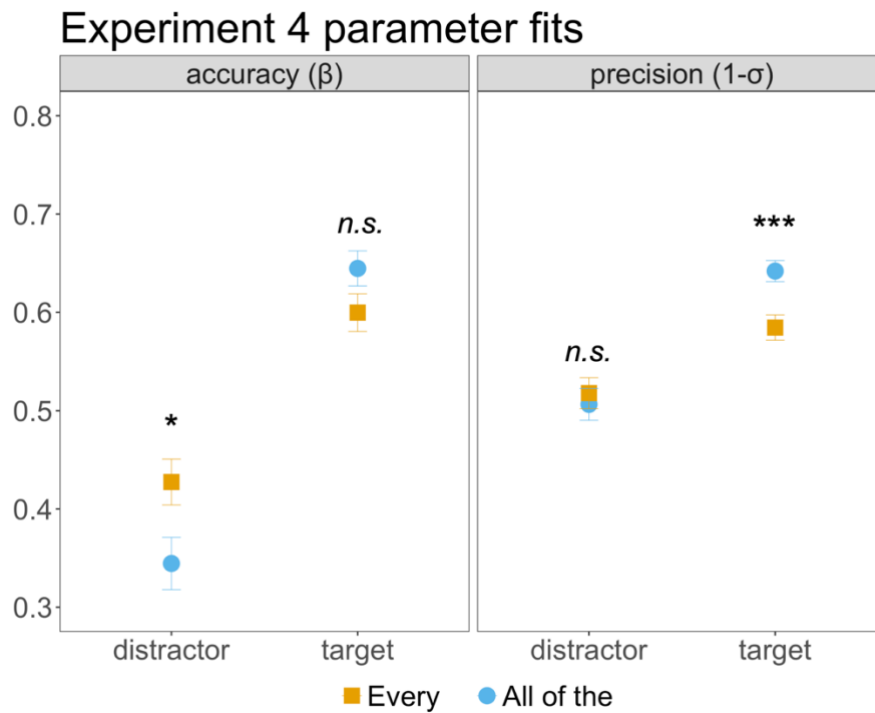


Figure 8: Parameter fits for participants' grouped responses on the cardinality portion of experiment 2. "Target" questions probed the restrictor set of the quantificational statement (e.g., *big dots* in *all of the big dots are blue*); "distractor" questions probed the complement of the restrictor set (e.g., *small* or *medium dots* in the above example).

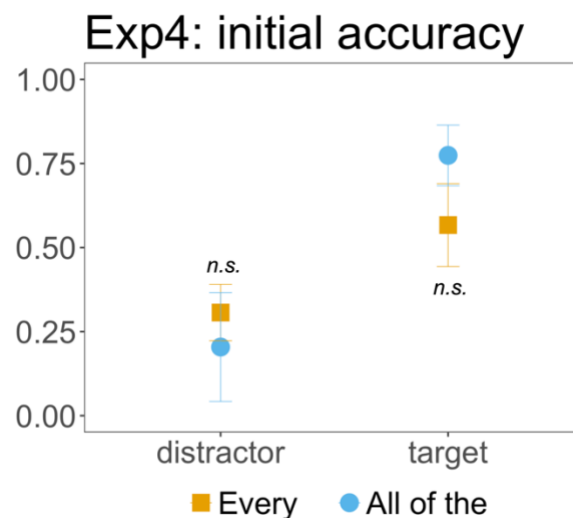


Figure 9: Accuracy (β) fits for each participant's first trial of each question type.

5. General discussion

We began with the question of whether the logical distinction between first- and second-order formats corresponds to a parallel psychological distinction. In order to look for evidence of this distinction, we probed participants' memory for various sets (by asking how well they could estimate those sets' cardinalities) following evaluation of different quantificational statements. 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 any group that the meaning highlights (e.g., "all of the big dots" highlights 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 predictions were substantially borne out. Independent of partitivity or plurality, statements with *there is a*, and *each* led to worse cardinality estimates for the relevant sets than statements with *most of the*, *all of the*, and *every*. We suggest that 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 – reflects a psychological analogue of the first-/second-order distinction. But there are other possible explanations, to which we now turn.

5.1 Distributivity

As we saw in section 1, sentences with *each* are always given distributive interpretations such that the predicate of an *each*-statement applies to the individuals in the domain (e.g., (30a) doesn't mean that the boys sang together). On the other hand, *all* is not always used in this way. The judgement generally reported in the literature (e.g., Beghelli & Stowell, 1997) is that the ease of accepting the distributive interpretation (in which there were as many singings as there were boys) is *each* > *every* > *all*, as in (30).

- (30) a. Each boy sang happy birthday (by himself / # in perfect harmony).
- b. Every boy sang happy birthday (by himself / ? in perfect harmony).
- c. All the boys sang happy birthday (by themselves / in perfect harmony).

The linking hypothesis between this apparent fact and our experimental task might look as follows: 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 *each*.

Notice that this linking hypothesis is about the word *each* being highly associated with distributive thoughts, not about the statements 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 *most of the dots* or *all of the dots* to be blue. Even when groups are implicated (as we suspect they are for *most*, *every*, and *all*) the property of blueness is distributed over the individual members. This raises a potential problem for an account 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 (30b) 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., Gil, 1995; Beghelli & Stowell, 1997; Tunstall, 1998; Winter, 2002; Champollion, 2017). 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 (31).

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

In particular, the *each* variant of (31) 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 (e.g., “it isn’t true that every dragon is dangerous”) (Szabolcsi, 2015).

The result from our experiments is that *every* and *all* pattern together to the exclusion of *each*. But in terms of being used distributively, *each* and *every* pattern largely together to the exclusion of *all*. Moreover, quantifiers like *most* can be and often are used distributively (e.g., in (32) the distributive reading is available and perhaps even preferred).

(32) Most boys sang happy birthday (by themselves > in perfect harmony).

As mentioned, *most* is a manifestly second-order quantifier. And in our experiments, *most* patterns more like *every* and *all* than like *each*. So the mere fact that a quantifier can be used or is often used in distributive contexts cannot explain our results.

One could maintain that the root cause of *each*’s obligatory distributivity is also what’s responsible for our results. For example, Beghelli & Stowell (1997) stipulate that *each* has a strong [+DISTRIBUTIVE] feature not present on *every* and *all*. One way or another, the content of 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 potentially 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 quantification in first-order logic 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-/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 would explain its mandatory distributivity as a direct consequence of its format. 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 sub-events to be distinct. Champollion (2015) differentiates *each* from the other universals by specifying associated presuppositions that provide the granularity over which they quantify (e.g., atoms for *each*; small sets for *all*). Both of these proposals include machinery that serves to highlight individuals, and as such, may be able to account for 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 individuals (Feigenson & Carey, 2003; Feigenson, Dehaene, & Spelke, 2004; Halberda, Sires, & Feigenson, 2006; Whitney & Leib, 2018, a.o.).

5.2 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 though, 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, 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.

5.3 The role of frequency

As one reviewer pointed out, uses of *each* are by far 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 summary statistics should lead participants to perform better following *each*-statements than *every*-/*all*-statements. Superior performance following *each*-statements could not as obviously be explained by excess processing costs incurred for *each*'s low frequency.

5.4 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 (i.e., the quantificational content is represented in no particular format; or in different formats across individuals). One reviewer rightly pointed

out that this does not exhaust the space of logical possibilities. 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.

Future work hopes to tease apart these hypotheses empirically. In our view, however, the idea of some quantifier meanings being representationally neutral and others being represented in particular (and universally-shared) formats seems to unnecessarily complicate the cognitive architecture. What would meanings have to be such that some of them are specified as the equivalence class of all truth-conditionally equivalent logical forms whereas others are specified in a single explicit logical form?

6. Conclusion

One tradition in Semantics treats expressions as names for things in the world: *rabbit* 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 (Davidson, 1967; Montague, 1973; Lewis, 1975; 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). 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, (1) better describes the mental representation of *each rabbit is furry* than (3) (and that the opposite is true for statements with *all* or *every*).

(1) $\forall x:Rx(Fx)$

(3) $R \subseteq F$

More generally, the proposed representations embrace the interface between linguistics and psychology in that they ground out in functions that are available elsewhere in cognition (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- and second-order representations are available to prelinguistic infants. We expect that they will be, and studies to test these predictions are currently underway.

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.

Acknowledgements

This work was supported by the National Science Foundation (Grant #1449815) and by the McDonnell Foundation (Grant on The Nature and Origins of the Human Capacity for Abstract Combinatorial Thought). We are grateful to Alexander Williams, Zoe Ovans, and three anonymous reviewers for valuable feedback on a previous draft. For helpful discussion throughout this project, we thank: Ellen Lau, Darko Odic, and audiences at MACSIM 2017, LSA 2019, CUNY 2019, Johns Hopkins University, and the University of Maryland.

Declarations of interest: none

References

- 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.
- Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. In *Philosophy, language, and artificial intelligence* (pp. 241–301). Springer.
- Beghelli, F., & Stowell, T. (1997). Distributivity and negation: The syntax of each and every. In *Ways of scope taking* (pp. 71–107). Springer.
- 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.
- Boolos, G. (1999). *Logic, logic, and logic*. Harvard University Press.
- Burr, D., & Ross, J. (2008). A visual sense of number. *Current Biology*, 18(6), 425–428.
- 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.
- Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision research*, 43(4), 393–404.

- Champollion, L. (2015). Stratified reference: the common core of distributivity, aspect, and measurement. *Theoretical Linguistics*, 41(3-4), 109-149.
- Champollion, L. (2017). Distributivity, collectivity and cumulativity. In *Companion to Semantics*. Wiley.
- Chierchia, G. (1998). Plurality of mass nouns and the notion of “semantic parameter”. In *Events and grammar* (pp. 53-103). Springer, Dordrecht.
- Church, A. (1941). *The Calculi of Lambda Conversion*. Princeton: Princeton University Press.
- Davidson, D. (1967). Truth and meaning. In *Philosophy, Language, and Artificial Intelligence* (pp. 93-111). Springer, Dordrecht.
- Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. OUP USA.
- Dowty, D. (1979). *Word Meaning and Montague Grammar*. Dordrecht, Holland: D.
- Dowty, D. (1987). Collective predicates, distributive predicates and all. In *Proceedings of the 3rd ESCOL* (pp. 97-115). Ohio: (Eastern States Conference on Linguistics), Ohio State University.
- Feigenson, L. and Carey, S. (2003) Tracking individuals via object-files: evidence from infants’ manual search. *Dev. Sci.* 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). 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 *Quantification in natural languages* (pp. 321-362). Springer, Dordrecht.
- Geurts, B., & Nouwen, R. (2007). 'At least' et al.: the semantics of scalar modifiers. *Language*, 533-559.
- 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., 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.
- Horty, J. (2007). *Frege on Definitions: A Case Study of Semantic Content*. Oxford: Oxford University Press.
- Im, H. Y., Zhong, S., & Halberda, J. (2016). Grouping by proximity and the visual impression of approximate number in random dot arrays. *Vision Research*, 126, 291–307.
- Ioup, G. (1975). Some universals for quantifier scope. *Syntax and semantics*, 4, 37–58.
- Knowlton, T., Yang, Y., Wong, A., Langfus, J., Pietroski, P., Lidz, J., & Halberda, J. (in prep). Two majority quantifiers in Cantonese bias distinct verification strategies.
- Krueger, L. E. (1984). Perceived numerosity: A comparison of magnitude production, magnitude estimation, and discrimination judgments. *Perception & 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* (No. 30). Oxford University Press.
- LaTerza, C. (2014). *Distributivity and plural anaphora* (Doctoral dissertation).
- Lewis, D. (1975). Languages and Language. In K. Gunderson (ed.) *Minnesota Studies in the Philosophy of Science*, volume 7. Minneapolis: University of Minnesota Press.
- Lidz, J., Pietroski, P., Halberda, J., & Hunter, T. (2011). Interface transparency and the psychosemantics of most. *Natural Language Semantics*, 19(3), 227–256.
- Link, G. (1983) The Logical Analysis of Plurals and Mass Terms. In *Meaning, Use, and Interpretation of Language*, ed. R. Bäuerle, C. Schwarze, and A. von Stechow, 303–323. De Gruyter.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. MIT Press. Cambridge, Massachusetts.
- Montague, R. (1973). The proper treatment of quantification in ordinary English. In *Approaches to natural language* (pp. 221–242). Springer, Dordrecht.

- 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.
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 1–9.
- Pietroski, P., Lidz, J., Hunter, T., & Halberda, J. (2009). The meaning of ‘most’: Semantics, numerosity and psychology. *Mind & 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.
- Pietroski, P. M. (2005). *Events and semantic architecture*. Oxford University Press.
- Pietroski, P. M. (2018). *Conjoining meanings: Semantics without truth values*. Oxford University Press.
- R Core Team. (2017). *R: A language and environment for statistical computing* [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Rescher, N. (1962). Plurality Quantification. *Journal of Symbolic Logic* 27: 373–4.
- Scha, R. 1984. Distributive, Collective, and Cumulative Quantification. In *Truth, Interpretation, and Information*, ed. Jeroen Groenendijk, Martin Stokhof, and Theo Janssen. Dordrecht: Foris.
- Schein, B. (1993). *Plurals and events* (Vol. 23). MIT Press.
- Schwarzschild, R. (1992). Types of plural individuals. *Linguistics and Philosophy*, 15(6), 641–475.
- Schwarzschild, R. (1996). *Pluralities* (Vol. 61). Springer Science & Business Media.
- 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.
- Szabolcsi, A. (2010). *Quantification*. Cambridge University Press.
- Szabolcsi, A. (2015). Varieties of quantification. In N. Riemer (Ed.), *The routledge handbook of semantics* (chap. 18). Routledge.

- Tichý, P. (1969). Intension in terms of Turing machines. *Studia logica*, 24(1), 7-21.
- Tomaszewicz, B. M. (2011). Verification strategies for two majority quantifiers in polish. In *Proceedings of sinn und bedeutung* (Vol. 15).
- Tunstall, S. (1998). The interpretation of quantifiers: Semantics & processing (Unpublished doctoral dissertation). Ph.D. dissertation, Department of Linguistics, University of Massachusetts at Amherst.
- Vendler, Z. (1962). Each and every, any and all. *Mind*, 71(282), 145–160.
- 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*. London: Routledge & Kegan Paul.
- Williams, A. (2015). *Arguments in syntax and semantics*. Cambridge University Press.
- Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual review of psychology*, 69, 105-129.
- Winter, Y. (2002). Atoms and sets: A characterization of semantic number. *Linguistic Inquiry*, 33(3), 493-505.