GLOSSA
psycholinguistics

eScholarship
University of California

Individuals versus ensembles and *each* versus *every*: Linguistic framing affects performance in a change detection task

Tyler Knowlton, MindCORE, University of Pennsylvania, US, tzknowlt@upenn.edu

Justin Halberda, Department of Psychological and Brain Sciences, Johns Hopkins University, US, halberda@jhu.edu

Paul Pietroski, Department of Philosophy, Rutgers University, US, paul.pietroski@rutgers.edu

Jeffrey Lidz, Department of Linguistics, University of Maryland, US, jlidz@umd.edu

Though *each* and *every* are both distributive universal quantifiers, a common theme in linguistic and psycholinguistic investigations into them has been that *each* is somehow more individualistic than *every*. We offer a novel explanation for this generalization: *each* has a first-order meaning which serves as an internalized instruction to cognition to build a thought that calls for representing the (restricted) domain as a series of individuals; by contrast, *every* has a second-order meaning which serves as an instruction to build a thought that calls for grouping the domain. In support of this view, we show that these distinct meanings invite the use of distinct verification strategies, using a novel paradigm. In two experiments, participants who had been asked to verify sentences like *each/every circle is green* were subsequently given a change detection task. Those who evaluated *each*-sentences were better able to detect the change, suggesting they encoded the individual circles' colors to a greater degree. Taken together with past work demonstrating that participants recall group properties after evaluating sentences with *every* better than after evaluating sentences with *each*, these results support the hypothesis that *each* and *every* call for treating the individuals that constitute their domain differently: as independent individuals (*each*) or as members of an ensemble collection (*every*). We situate our findings within a conception of linguistic meanings as instructions for thought building, on which the format of the resulting thought has consequences for how meanings interface with non-linguistic cognition.

1. Introduction

It's well known that linguistic details can influence the thoughts that listeners construct in response to sentences that are equivalent in terms of their objective content. Given the same state of the world, logically equivalent sentences can highlight different aspects of it. To take one example, the verbs *chase* and *flee* can be used to describe the same event, but they highlight different perspectives (Gleitman, 1990). To borrow another example, Duchon, Dunegan, and Barton (1989) asked engineers to decide whether a hypothetical research and development team should be given $100,000. Half of the participants were told of the team's past success rate, as in (1a); the other half was told of the team's failure rate, as in (1b).

(1)   a.   Of the projects undertaken by the team, 30 of the last 50 have been successful.
        b.   Of the projects undertaken by the team, 20 of the last 50 have been unsuccessful.

Participants were more likely to agree to fund the hypothetical team given the framing in (1a), despite the historical success rate being 60% in either case (for similar examples, see Geurts, 2013; Kahneman & Tversky, 1979; Tversky & Kahneman, 1981).

In light of such examples, one might wonder how deep linguistic framing effects run. To what extent will distinct but logically equivalent descriptions of a situation give rise to psychologically distinct thoughts? Given a mentalistic view of meaning, one expects linguistic framing to be ubiquitous. On such a view, meanings relate to non-linguistic cognitive systems and concepts in a way analogous to how pronunciations relate to motor-planning systems and articulators (e.g., Chomsky, 1964; Jackendoff, 1983; Pietroski, 2018). And if sentence meanings serve as instructions for how to build thoughts, then distinct sentences that correspond to logically equivalent descriptions of situations can still have distinct meanings that lead to assembly of importantly different thoughts.

Here, we focus on one such case: the quantificational determiners *each* and *every*. Compared to the *chase* versus *flee* and *successful* versus *unsuccessful* examples above, *each* and *every* initially seem less likely to give rise to a framing effect. After all, whatever these words mean, the important similarities between them are clear. Both are universal (as opposed to existential or proportional) quantifiers, and both are distributive (in that if *each/every F is G* then *every single one of the Fs is G*). So it is guaranteed that in any situation, *each F is G* if and only if *every F is G*. One expects competent speakers to know this equivalence and to be able to infer from *each F is G* to *every F is G,* and vice versa. In contrast, though *chase* and *flee* are related, they are not themselves logically equivalent (and the inference from *F chased G* to *F fled G* is not guaranteed). So while sentences with *chase* and *flee* give rise to distinct thoughts, one might attribute differences between these thoughts to different associations speakers have with the relevant lexical items or to the different syntactic roles played by their arguments. Either way, it should come as less of a surprise that such expressions are understood differently. Here, though, we argue that *each* and *every* are also

understood differently: despite being logically equivalent, they are used to assemble thoughts with distinct conceptual constituents.

The conceptual constituents at issue are representations of groups and individuals. Consider various ways of formally specifying the informational contribution that *each* or *every* make to a sentence like (2a). In particular, consider the specifications in (2b-c), which differ in which group(s) they explicitly encode (these representations expand on the hypotheses originally introduced in Knowlton, 2021, and Knowlton et al., 2022).

(2)   a.   Each/Every circle is green.

        b. $\forall x{:}\text{Circle}(x)[\text{Green}(x)]$

           $\approx$ any thing$_x$ that is a circle is such that it$_x$ is green

        c. $\text{TheX}{:}\text{Circle}(X)[\forall x{:}X(x)[\text{Green}(x)]]$

           $\approx$ the circles$_X$ are such that each one of them$_X$ is green

        d. $\forall x{:}\text{Circle}(x)[\text{TheY}{:}\text{Green}(Y)[Y(x)]]$

           $\approx$ any thing$_x$ that is a circle is such that it$_x$ is one of the green-things$_Y$

        e. $\text{TheX}{:}\text{Circle}(X) \subseteq \text{TheY}{:}\text{Green}(Y)$

           $\approx$ the circles$_X$ are included in the green-things$_Y$

All four specifications encode the universal quantificational content expressed by *each* and *every*. But (2b-e) differ with regard to whether they include constituents that represent the circles and the green things as such.[1] While (2b) includes a predicate that applies to circles ('Circle(x)') and a predicate that applies to green things ('Green(x)'), it has no constituent that explicitly represents the things that either predicate applies to. It expresses a thought that essentially abbreviates a conjunction of color-claims about particular circles, with a conjunct for each circle (circle$_1$ is green & circle$_2$ is green & circle$_3$...). By contrast, (2e) represents both groups; indeed, it represents them *as* the circles and the green things ('TheX:Circle(X)' and 'TheY:Green(Y)'). Specifications (2c) and (2d) treat 'circle' and 'green' asymmetrically: (2c) represents the circles, but not the green things; (2d) represents the green things, but not the circles. With these options in mind, one can ask which one is a better description of how speakers of English understand *each*, *every*, and sentences like (2a).

Prior work in this vein has defended the claim that *every* is understood in a way that calls for grouping the domain of quantification (e.g., the circles in *every circle is green*), as suggested by the semi-second-order specification in (2c). In particular, Knowlton et al. (2022) draw

---

[1] In (2b-e), expressions like 'TheX:Circle(X)' are intended to be understood as "the circles". This could be elaborated in many ways, including with an iota operator: '$\iota X{:}\forall x(Xx \equiv \text{Circle}(x))$'. That is, 'TheX:Circle(X)' is shorthand for "the things X such that for each thing x, x is one of them (the Xs) if and only if x is a circle". Alternatively, 'TheX:Circle(X)' can be taken to indicate the set of circles: '{x: x is a circle}'. The important point, for present purposes, is just that the second-order expression 'TheX:Circle(X)' is a constituent that represents the circles as such, whereas the predicate 'Circle(x)' is not, and neither is the quantifier '$\forall x{:}\text{Circle}(x)$'.

this conclusion on the grounds that being asked to verify sentences with *every* encourages participants to represent the domain using their non-linguistic cognitive system for ensemble representation.[2] In this paper, we defend the complementary claim that *each* is understood in a way that includes no notion of grouping the domain of quantification, as suggested by the purely first-order specification in (2b). We do so by providing evidence suggesting that verifying sentences with *each* encourages participants to represent the domain using their non-linguistic system for object individuation. The resulting claim is that *each*'s meaning is completely first-order, in contrast to *every*'s meaning, which has a second-order component. We further claim that this difference in meaning invites the recruitment of different extralinguistic cognitive systems: object-files and ensembles. But episodes of invoking these non-linguistic representational systems are downstream consequences of how the distinct linguistic expressions are understood (i.e., downstream consequences of the meanings in (2b-c)). We are not proposing that non-linguistic representations like object-files and ensembles are in any way part of the meaning of these expressions. Rather, the idea is that *each* and *every* have as their meanings mental representations that are logically equivalent but psychologically distinct in a way that explains their different effects on other aspects of cognition.

1.1 reviews prior linguistic and psycholinguistic work on *each* and *every*, which, in our view, initially motivates considering (2b-c) as hypotheses about their meanings. 1.2 discusses these hypotheses in more detail, including the proposed relationship between the hypothesized linguistic representations (those in (2b-c)) and a pair of well-studied non-linguistic representations (object-files and ensembles). 1.3 motivates the current experiments – a change detection task following a sentence verification task – which are presented in Sections 2 and 3. Finally, Section 4 addresses whether alternative views about the meanings of *each* and *every* would have predicted the same results, and concludes.

1.1 Linguistic and psycholinguistic background

A unifying theme in prior work on universal quantifiers might be put as follows: *each* highlights individuals a bit more than *every* does, whereas *every* is slightly friendlier to groups than *each* (though far less friendly to groups than *all*). As Vendler (1962) put it, drawing on historical evidence: "*every* comes from *ever each*, thus originally it served to sum up the distribution characteristic of *each*. In this sense, *every* is between *each* and *all*" (p. 149). This intuition – that *each* is more individualistic in some sense – is our leading idea. The question is in what sense *each* is more individualistic. Instead of taking the usual tack of diagnosing it as a peripheral fact, accommodated by positing slight syntactic differences in two semantically indiscernible universal quantifiers, we think this relatively subtle difference between *each* and *every* is an important

---

[2] And, on similar grounds, Knowlton, Pietroski, et al. (2021) consider and reject (2d-e) as meanings for *every*.

phenomenon that reflects a theoretically important contrast in the mental representations that these expressions are used to access. In this section, we review some of the previously reported linguistic intuitions that we think initially motivate considering representations like those in (2b-c), but which were not initially reported with those representations in mind.

Landman (2003) notes some delicate contrasts between *each* and *every* that are presumably symptoms of a difference in meaning. For example, compared with *each, every* combines more comfortably with collective verbs like *combine* as in (3).

(3)   a. #In this class I try to combine each theory of plurality.
        b.   In this class I try to combine every theory of plurality.

And as seen in (4), *every NP* is better than *each NP* for describing a group.

(4)   a. #The press is each person who writes about the news.
        b.   The press is every person who writes about the news.

These contrasts suggest that *every* supports grouping its internal argument – *theory of plurality* in (3), *person who writes about the news* in (4) – in a way that *each* does not (Landman captures this formally by positing that *every NP* can shift between a quantificational interpretation and a definite interpretation).

Moreover, consider (5); see Surányi (2003) and Szabolcsi (2010).

(5)   a.   Determine whether each number in this list is odd: <1, 3, 4>.
        b.   Determine whether every number in this list is odd: <1, 3, 4>.

Given *each*, one can respond felicitously without saying "yes" or "no," but instead saying "1 is odd; 3 is odd; but 4 is not." This sort of response, a pair-list reading (May, 1985), considers each element of the list in turn. So for each number on the list, one says whether or not it is odd. Given *every*, the same sort of pair-list response seems less natural (though not impossible). One is inclined to reply about the whole list at once, and then perhaps elaborate by saying something like "No, every number on that list isn't odd" or "No, 4 isn't odd."

A similar point applies to examples like (6), modified from Beghelli (1997).

(6)   a.   Q: Which book did you loan to each student?
             A: *Frankenstein* to Frank, *Persuasion* to Paula, and *Dune* to Dani.

        b.   Q: Which book did you loan to every student?
             A: #*Frankenstein* to Frank, *Persuasion* to Paula, and *Dune* to Dani.

While the pair-list answer provides a possible response to the *each*-question in (6a), it would be a comparatively infelicitous response to the *every*-question in (6b). Again, it seems that *every NP* pushes one to consider the plurality of NPs, whereas *each NP* pushes one to consider the individual things that satisfy *NP*.

Relatedly, *each* seems to resist "generic" interpretations in cases where *every* seems to invite them. For instance, Beghelli and Stowell (1997) discuss examples similar to (7), which they attribute to Gil (1992).

(7)   After a lifetime of investigation, Suzie concluded that
    a.   every galaxy has a black hole at its center.
    b.   #each galaxy has a black hole at its center.

One way to understand these sorts of examples (though somewhat distinct from Beghelli and Stowell's original presentation) is that the *every* variant in (7a) suggests a domain of quantification that goes well beyond the particular galaxies Suzie studied. The *each* version in (7b) does not. Instead, (7b) sounds anomalous because it implies that Suzie's conclusion generalizes no further than the particular galaxies she studied. Put another way, a truly universal generalization can more easily be stated over *every* galaxy than over *each* galaxy. Conversely, *each* feels more natural in contexts where the claim in question is not intended to extend beyond the local domain. For example, the acceptability of *each* and *every* in (8) is reversed.

(8)   Suzie just discovered four new galaxies in a distant region of space and concluded that
    a.   #every galaxy has a black hole at its center.
    b.   each galaxy has a black hole at its center.

That is, *every* seems to invite a generalization that is far too broad in (8a). Given the context, what's at issue is not a universal generalization but details about the particular four galaxies Suzie discovered. Whereas *each* is compatible with this individualistic thought, *every* resists it.

In the same vein, (9a) is naturally understood as a component of a drink recipe, whereas (9b) is more naturally understood as a claim about some particular cocktails in need of garnishes (Knowlton et al., 2023).

(9)   a.   Every Old Fashioned needs an orange peel.
    b.   Each Old Fashioned needs an orange peel.

The judgments above are subtle. But Tunstall (1998) provides some confirming evidence. In one experiment, participants were presented with short stories and a choice between *each* or *every* (e.g., "When Max visited the store he wrote down on his notepad what {each/every} employee was wearing"). The story leading up to the choice either highlighted differences between the things quantified over (e.g., "On Monday, the deli clerk had on a striped shirt and the cashier in the express lane had on a floral shirt.") or similarities between them (e.g., "Anyone with long hair had to put it up in a pony tail."). Participants were more likely to choose *each* in the former case and more likely to choose *every* in the latter.

Potentially relevant evidence also comes from priming experiments. Feiman and Snedeker (2016) find that a given interpretation of scopally ambiguous sentences with *each* primed that same interpretation in other sentences with *each*, but not in other sentences with *every*. Likewise, a given interpretation of sentences with *every* primed that same interpretation of other sentences with *every* but not with *each*. The same results were found even when the verbs were changed, suggesting the lack of *each-every* priming was not merely due to altering a single lexical item. At the same time, they found that number words primed other number words (e.g., *three* primes *four*). Assuming priming reflects similarity of meaning, these results can be taken to suggest that *three* and *four* have more similar meanings than *each* and *every*, despite the latter but not the former sharing truth-conditions.

More directly relevant here, Knowlton et al. (2022) showed success at measuring a cognitive reflection of the group-friendly character of sentences containing *every* to a greater degree than for otherwise equivalent sentences containing *each*. Participants were first presented with a sentence like {*each/every*} *big dot is blue* along with a picture of different sized and different colored circles. Their initial task was to evaluate the sentence (i.e., to answer "true" or "false"). After they answered, participants were then asked to estimate the cardinality of a random group of dots (e.g., *how many big dots were there*?), which is fundamentally a group property. Participants were better able to recall the number of circles after evaluating an *every*-sentence than an *each*-sentence. These effects held even within participants, suggesting that changing *each* to *every* encouraged them to treat the exact same scene in a different way. Namely, *each* was less likely than *every* to lead participants to group the circles into a collection whose cardinality could be estimated.

None of the above results are categorical. This suggests to us that they don't deserve a grammatical explanation. Still, these subtle facts suggest to us that while *each* and *every* are importantly similar, *each* is somehow more individualistic and *every* is somehow friendlier to groups. The challenge is how to account for this (subtle and non-categorical) difference while retaining the (more obvious) fact that *each* and *every* are both distributive universal quantifiers. And though various proposals exist for capturing a subset of the facts discussed above, none have yet provided a unified explanation of why *each* and *every* behave differently. In our view, considering what is known about extralinguistic systems for representing individuals-as-such versus individuals-as-group-members can help clarify the sense in which *each* and *every* differ.

1.2 A psycho-logical explanation

In part hoping to give the above differences a unified explanation, we propose that the meanings of *each* and *every* are instructions to assemble formally distinct mental representations along the lines of (2b) and (2c), repeated here:

(2) a. Each/Every circle is green.

    b. $\forall x{:}Circle(x)[Green(x)]$ EACH

        $\approx$ any thing$_x$ that is a circle is such that it$_x$ is green

    c. $TheX{:}Circle(X)[\forall x{:}X(x)[Green(x)]]$ EVERY

        $\approx$ the circles$_x$ are such that each one of them$_x$ is green

On this view, *each circle is green* is understood like a suitably exhaustive conjunction of individual color-claims about particular circles (e.g., "circle$_1$ is green & circle$_2$ is green & circle$_3$ is green & …"). The noun (phrase) with which *each* combines supplies a predicate that restricts the domain of quantification – any thing *that is a circle* as opposed to any thing *in the universe* – but does not correspond in any way to a grouping of this restricted domain. That is, the thought in (2b) has no part that represents the circles as such. In contrast, *every* is understood in a way that calls for mentally grouping the things specified via the noun (phrase) with which it combines, as (2c) makes explicit. Here, *every circle is green* can be thought of as roughly meaning "the circles are such that each one of them is green". In both (2b) and (2c) then, '$\forall x$' reflects the universal and distributive character of the quantifier; while *each* differs from *every*, these expressions are also fundamentally similar. In both cases, truth requires that the predicate given by the verb phrase (e.g., *is green*) applies to each and every member of the restricted domain. The difference is that the thought in (2c) calls for grouping the things that constitute the restricted domain (e.g., *the circles*), whereas the thought in (2b) does not (e.g., *individuals that are circles*).

    The first-order versus second-order distinction at issue between the forms in (2b) and (2c) is similar to distinctions that have previously been considered in investigations of quantifiers and their computational complexity (for review, see Szymanik, 2016). For example, building on the pioneering work of van Benthem (1986), Szymanik and Zajenkowski (2010) argue that the quantifiers that can be modeled with finite-state automata (e.g., *each*, *every*, and *all*) encourage distinct verification strategies from quantifiers like *most* that call for more powerful models (e.g., push-down automata). Other researchers have argued for different classifications (e.g., Clark & Grossman, 2007; McMillan et al., 2005; Olm et al., 2014). But in all such proposals, *each* and *every* fall under the same category, as both can be modeled with finite-state automata and both can be expressed in purely first-order terms. In our view, while both *each* and *every* can be modeled (by theorists) in the first-order way given in (2b), only *each* actually is specified that way in the minds of speakers.

    For us, the forms displayed in (2b) and (2c) are posited as mental representations that can be described as "psycho-logical forms".[3] The idea is that understanding a sentence is a matter of connecting its pronunciation (via its syntactic structure) with a certain *structured representation*

---

[3] As far as we know, the term "psycho-logical form" was coined by Soames (1987), in a review of Hornstein's (1984) *Logic as grammar*.

of the relevant environment. This representation/meaning may have a coarse-grained content that is shared with other representations (e.g., truth-conditions). But internal differences can matter, even if they do not mark mind-independent distinctions. This mentalistic perspective on linguistic meaning thus invites questions about how understanding distinct but content-equivalent expressions differently might encourage different non-linguistic representations. In this particular case, two well-studied psychological systems seem potentially relevant. The first is the object-file system (e.g., Green & Quilty-Dunn, 2021; Kahneman & Treisman, 1984; Kahneman et al., 1992), a system for representing individuals and their properties. The second is the ensemble representation system (e.g., Ariely, 2001; Whitney & Yamanashi Leib, 2018), a way of representing individuals as members of a collection.

An *object-file* is essentially an index of a particular individuated object that serves to anchor a list of individual properties (e.g., the object's size, its color, and its location in space). Given the high-fidelity nature of these representations, they are subject to a strict working memory limit: only 3-4 object-files can be simultaneously represented (e.g., Feigenson & Carey, 2005; Vogel et al., 2001). To get around this limit, a separate system of *ensemble* representation allows for representing a collection of individuated objects simultaneously. Ensembles accomplish this by abstracting away from individual properties and encoding the collection in terms of summary statistics (e.g., the average size of the individuals, their cardinality, and their center of mass). Both of these cognitive systems are operative in humans as early as infancy (for helpful reviews, see Carey, 2009; Feigenson et al., 2004). Much of the work on object-files and ensembles has focused on the visual domain (including the experiments reported below). It is possible to attend to three visual items either as an ensemble group (e.g., Alvarez, 2011; Alvarez & Oliva, 2008; Ariely, 2001; Chong & Treisman, 2003; Haberman & Whitney, 2011, 2012; Sweeny et al., 2015; Ward et al., 2016; Whitney & Yamanashi Leib, 2018) or as individuals (e.g., Kahneman & Treisman, 1984; Kahneman et al., 1992; Pylyshyn 2001; Pylyshyn & Storm, 1988; Xu & Carey, 1996; Xu & Chun, 2009). But these systems of representation are modality-neutral, as studies have extended them to visual events (Wood & Spelke, 2005), auditory beeps (Izard et al., 2009; Kanjlia et al., 2021; Piazza et al., 2013), and touches on the skin (Plaisier et al., 2009; Riggs et al., 2006). The fundamental distinction between individuals and ensembles is thus thought to be general throughout cognition.

The differences between the psycho-logical forms displayed in (2b) and (2c) suggest a hypothesis about how the meanings of *each* and *every* might be related to these ancillary non-linguistic psychological systems. Namely, the second-order component of *every*'s meaning – 'TheX:Circle(X)' – might invite forming an ensemble representation of the circles (or whatever things constitute the domain of quantification) in a way that the meaning of *each* – which lacks any constituent that represents the circles as such – will not. Instead, the purely first-order *each* might be expected to trigger purely individualistic object-file representations of the things

constituting the domain. To be clear, the idea is not that ensembles or object-files are part of the lexical entries given in (2b-c), nor is the idea that the meanings of *each* and *every* are in any sense parasitic on the existence of these two non-linguistic representational systems. One can easily imagine a mind that lacks object-files or ensembles but can nonetheless think thoughts like those indicated by (2b) and (2c) and can thus be said to know the meanings of *each* and *every*. Rather, the proposal is two-fold. First, *each* and *every* have as their meanings mental representations that differ in whether they contain a constituent corresponding to a second-order grouping of their internal argument (the restricted domain). Second, as a downstream consequence of this formal difference in meaning, *each* and *every* invite and often trigger the deployment of the different non-linguistic systems discussed above.

This proposal – both the meanings of *each* and *every* and the relation between those meanings and ancillary cognitive systems for representing object-files and ensembles – can potentially help explain the differences noted in 1.1. Some of the contrasts fall out of the formal difference between the psycho-logical forms in (2b) and (2c), without making any recourse to the object-file/ensemble distinction. For example, sentences with *each* or *every* don't prime each other, despite being logically equivalent, because their meanings are instructions to assemble thoughts with different contents. Likewise, the domain of quantification can be referenced as a group more easily in the *every*-variants than in the *each*-variants of (3-6) from 1.1, because that group is explicitly encoded in the representation only when *every* is used to indicate universal quantification.

Other contrasts can be explained given known properties of the associated object-file and ensemble systems. Take the "generic" cases in (7-9), where *every*-claims are more naturally understood as being intended to project beyond the locally established domain (*each dog barks* versus *every dog barks*). In these sorts of sentences, the domain quantified over with *every* is taken to be larger than the domain quantified over with *each*. If *each* invites representing the things constituting the domain as a series of object-files whereas *every* invites representing domain entities as members of an ensemble collection, then speakers' preference for using *every* for larger domains follows from the differential working memory limit on object-files and ensembles. Moreover, ensembles allow for representing large groups of things in a way that supports generalization. In particular, they license predictions about new group members by virtue of how they encode information, and they allow for vague group boundaries (Knowlton, 2021; Knowlton et al., 2023). So *every*'s apparent friendliness to genericity and *each*'s resistance to it can be explained not by the psycho-logical forms themselves, but by the ancillary extralinguistic systems they trigger. Indeed, this explanation accords with the fact that the contrast is not categorical: speakers can use *each* with large domains (*each star will burn out one day*) and they can use *every* when they have no intention of generalizing at all (*eat every veggie on your plate*).

The non-deterministic link between (2b) and (2c) and object-files and ensembles is thus an important feature of the proposal.[4]

More generally, we might put the point as follows: *every* feels somewhat friendly to groups in part because its meaning is an instruction to build a thought like (2c), which explicitly calls for grouping the domain (captured, in our notation, via the second-order variable, 'X'). This grouping, in turn, invites and often triggers the psychological system for ensemble representation, which has additional downstream consequences for language use. On the other hand, *each* feels highly individualistic because its meaning eventuates in a thought like (2b), which does not have a constituent that represents the domain as a group. Not representing the things constituting the domain as a group, but nonetheless universally quantifying over them, invites and often triggers the psychological system for object-file representation, which has downstream consequences of its own. But this contrast in friendliness to groups is non-categorical for two reasons. First, the meanings of both *each* and *every* are distributive (cp. the group-friendliness exhibited by plural collective nouns, as in *the army of frogs gathered by the pond then surrounded the flock of geese*). Second, since the invitation to represent the individuals that constitute the domain with the object-file or ensemble system is not deterministic, the downstream effects of deploying these systems is not always observed.

As noted, there have been other attempts at capturing the distinctions discussed in 1.1 (albeit in less overtly mentalistic frameworks). We return to some such alternatives in Section 4. In the meantime, we aim to provide new evidence for the two-part hypothesis offered in this section: the meanings of *each* and *every* are instructions to assemble formally distinct thoughts, and these thoughts naturally interface with different extra-linguistic cognitive systems.

1.3 Motivating the current experiments

As noted, previous work has focused on demonstrating that *every* promotes grouping its domain by asking whether participants recall group properties like cardinality and center of mass. For example, Knowlton et al. (2022) reported that when participants were asked overtly group-based questions like "how many big circles were there?", they performed better if they had first evaluated a sentence like *every circle is green* instead of *each circle is green*. Here, we seek the

---

[4] The exact nature and strength of the link between linguistic meanings like (2b-c) and non-linguistic systems like object-files and ensembles remains an open empirical question. Results like the ones presented below suggest it is at least strong enough to bias speakers to use certain strategies when evaluating thoughts for truth against the world. And, if the explanation of judgments like (7-9) does lie with details of these non-linguistic systems, then the link must be operative even outside of selecting verification strategies. Likewise, if children do use this sort of information during language acquisition, the link must be stronger still. But it is likely not a deterministic link, given that speakers can and do use *each* with large domains in everyday life (including in situations where they plausibly are not representing the domain as a series of object-files).

complement: to show that *each* promotes representation of individuals and their properties and thus leads to superior performance when overtly individualistic memory questions are posed. That is, while past work has argued that evaluating a sentence like *every circle is green* encourages creating an ensemble representation of the circles and thus encoding their summary statistics (with *each* as a control), the present study aims to find evidence that a sentence like *each circle is green* invites participants to create individual object-file representations of the circles, and thus precisely encode individual circles' properties (with *every* as a control).

If the view outlined in 1.2 is correct, the meaning of *each circle is green* is an instruction to build a representation that has no part corresponding to the circles. It is a purely first-order thought invoking only notions of objects and properties predicated of them. In evaluating this kind of thought for truth against some particular circles, participants are predicted to deploy individual object-file representations of the contextually relevant circles and determine whether the property of being green applies to each. Consequently, asking someone whether *each circle is green* is true will bias them to encode each circle's individual properties (e.g., the particular hue and position of each circle) but not necessarily group properties (explaining why participants are worse at follow-up questions probing cardinality or center of mass).

On the other hand, the meaning of *every circle is green* is an instruction to build a representation that has a constituent that treats the circles as a group. So in evaluating this sort of thought, speakers are predicted to be biased to create an ensemble representation of the circles and determine whether the property of being green applies to each member of this collection (note that it needs to apply to each member in the collection, not to the collection collectively, since *every* is still a distributive universal quantifier). How exactly participants go about this latter step of determining whether the predicate applies distributively to the members of the ensemble is less important for present purposes. There are numerous possibilities, whose availability will likely depend on the predicate in question. For *every circle is green*, participants might rely on summary statistics, like the range of hues in the collection (if the range of hues extends beyond the boundaries of being green, then at least one circle isn't green). Alternatively, they might create an ensemble, then subsequently individuate that ensemble's members and check whether the predicate applies to each one. In any case, the important claim here is that *every* will push participants toward an initial ensemble coding of the things quantified over. The crucial prediction is thus that participants will be more likely to represent the domain as a series of object-files – and accordingly will have better memory for individual hues – given *each* than given *every*.

To test this prediction, participants in the experiments below completed a joint sentence verification and change detection task. They were first shown three circles with different hues (e.g., **Figure 1**) and were either asked to evaluate sentences like *each circle is green* or *every circle is green*. After participants responded, the circles disappeared. Following a brief grey screen,

the circles returned in the same spatial locations. Sometimes, a single circle's hue was altered. Participants were then asked to evaluate a second sentence, which invited circle individuation: *one circle changed its color*. We predicted superior change detection ability for those participants in the *each* condition compared to those in the *every* condition, despite being shown the same displays and offering the same responses to the initial question.
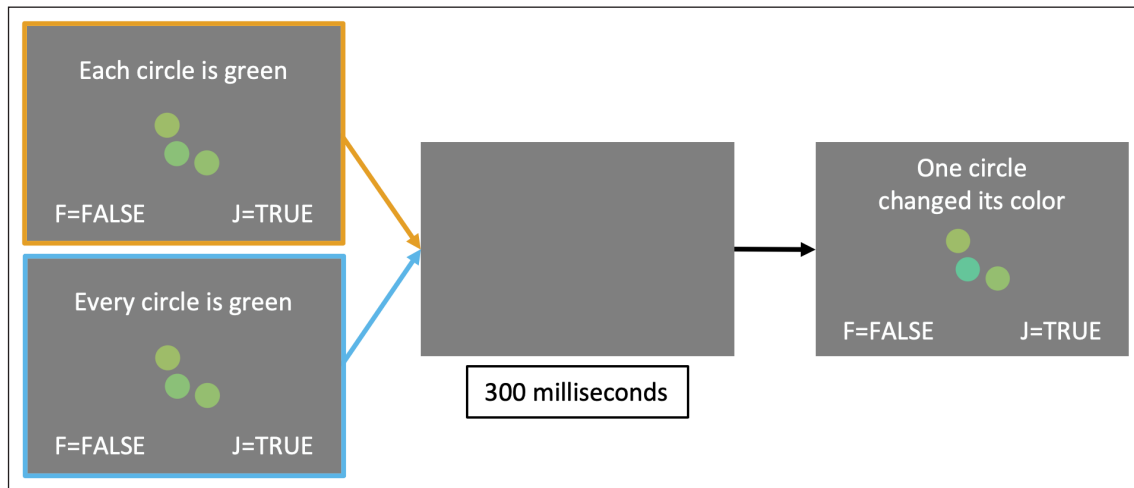
Figure 1: Trial structure of the experiments. In this example trial, the correct answer to the follow-up question is TRUE, as the middle circle changed its hue. Quantifier used in the initial sentence was manipulated between subjects.

This prediction relies on the idea that a small number of items can be represented either as independent object-files or as a single ensemble. This is perhaps counterintuitive, as small numbers of items generally trigger object-file representations whereas large numbers of items generally trigger ensemble representations (Feigenson et al., 2004). To take one illustrative example, infants can reliably distinguish 8 from 16 objects (Xu & Spelke, 2000) but fail to distinguish 1 from 4 objects (Feigenson & Carey, 2005). In the 8 versus 16 case, the large numbers encourage treating the groups of objects as ensembles, and enumerating and comparing their approximate cardinalities. But in the 1 versus 4 case, the smaller numbers encourage treating the objects as independent object-files, which pushes infants beyond their working memory capacity and leads to failure (in contrast, infants can reliably distinguish 1 versus 3 objects when treating those objects as independent individuals; see Feigenson & Carey, 2003). This large versus small distinction can even be found in identical experimental setups: 6-month-old infants successfully distinguished 8 versus 4 actions but failed to distinguish 4 versus 2 actions (Wood & Spelke, 2005). In other words, even when representing four things as an ensemble would lead to superior performance (by getting around working memory constraints), the context of presentation involving only small numbers encouraged treating them as independent object-files instead.

Here, *every* is predicted to lead participants to treat the relevant objects as members of an ensemble despite the fact that (i) they are present in small numbers and (ii) doing so would lead to inferior performance on the subsequent change detection task. Showing superior performance for a follow-up memory question given *each* is important, as past work on universal quantifiers and verification strategies has only shown superior performance for *every* (when the question probes a group property, like cardinality). As such, past work leaves open the possibility that *each* always leads to worse performance on memory questions (perhaps because its relative infrequency requires extra cognitive effort that might otherwise be spent encoding information like the domain's cardinality in memory). This complementary finding would help rule out such low-level explanations.

2. Experiment 1: Constant difficulty

2.1 Participants

43 participants were recruited on Amazon Mechanical Turk. All gave informed consent and passed an English-screener prior to participating in the actual task. One participant was removed from further analysis for having response times longer than three standard deviations above the mean. Six participants were removed from further analysis for achieving below 55% "correct" on the sentence verification portion of the task (as determined by responses to a color naming task from Bae et al., 2015; see below). This left 36 participants.

2.2 Stimuli and procedure

The task consisted of 84 trials of sentence verification followed by change detection. On each trial, participants first read a quantificational sentence. Six different sentences were used: {*each/ every*} *circle is* {*green/blue/orange*}. The color used varied randomly from trial to trial, to provide variety to the task. Quantifier was manipulated between subjects, with half of the participants evaluating a series of *each*-sentences and half of the participants evaluating a series of *every*-sentences. We opted for a between-subjects design given that past work probing differences between *each* and *every* revealed order effects. In particular, participants in a within-subjects experiment showed better cardinality estimation accuracy following *every* than *each*, but the effect was far more pronounced among those participants who first completed the *each* block (Knowlton, 2021). The current between-subjects design allows us to sidestep the potential for similar competition between conflicting linguistic framing (*each* priming an individual-based strategy and *every* priming a group-based one).

Each sentence was presented alongside a display of three colored circles on a grey background. Circle colors were randomly selected from an independently normed color wheel with a constant luminance and 180 hues (Bae et al., 2015). The wheel was designed such that each hue is equally

far apart from its neighbors in CIELAB color space. Based on Bae et al.'s norming data – which determined the hues that most participants accept as green, blue, and orange – sentences were designed to be true with respect to the display on half of the trials.

In particular, a hue was considered a member of a given color category if it was the modal response when adult participants were asked to name it. In a trial intended to be true, all three hues were randomly drawn from this empirically determined color distribution. By design, this led to borderline cases and left room for reasonable participants to disagree about whether a particular sentence is a good description of the corresponding display.

Participants were given as long as they wanted to inspect the circles and decide whether they agreed with the sentence. When they were ready to render their judgement, they did so by pressing "J" or "F" on their keyboard. The display then disappeared for 300 milliseconds, leaving a grey screen. Then three circles reappeared in the same spatial positions, alongside the text *one circle changed its color*. Participants again had unlimited time to evaluate this statement as true or false by pressing "J" or "F".

On half of the trials, no color change occurred. The original three circles reappeared on-screen following the 300 millisecond delay. On the other half of the trials, one randomly-selected circle changed its hue. In these cases, a new color was sampled from a Gaussian distribution over the hues, with a mean of the original hue and a standard deviation of 17. Selection of the original hue was not permitted. This method of random generation meant that the new hue often fell within the same color category as the original hue, making many of the change detection trials reasonably difficult. Participants were initially told that some of the changes could be small (and given an example of a within-category change), to ensure that they interpreted *one circle changed its color* to mean that a circle changed hue, not that a circle changed color category.

2.3 Results

Given the empirically determined color categories used to generate trials, participants correctly evaluated 67% of the *each*-statements and 69.7% of the *every*-statements. As noted above, many of the hues used straddled the border between two categories, so the seemingly low percent "correct" on this portion of the task in both conditions likely reflects disagreement between participants about, for example, which hues count as green. This was by design, to keep the task difficult and make participants think the task was about color categories, as opposed to the linguistic framing of the initial sentence.

Prior to rendering their judgments, participants viewed the displays for an average of 1767 milliseconds in the *every* condition and an average of 1627 milliseconds in the *each* condition. This difference in reaction time was not significant ($t_{31.67} = 0.71, p = .486$; fractional degrees of freedom result from Welch's t-test, which was used throughout).

On the change detection portion of the task, participants who had evaluated an *each*-sentence were more accurate than participants who had evaluated an *every*-sentence ($t_{33.97} = 2.33$, $p < .05$). Their accuracy on this portion of the task is plotted in **Figure 2**.
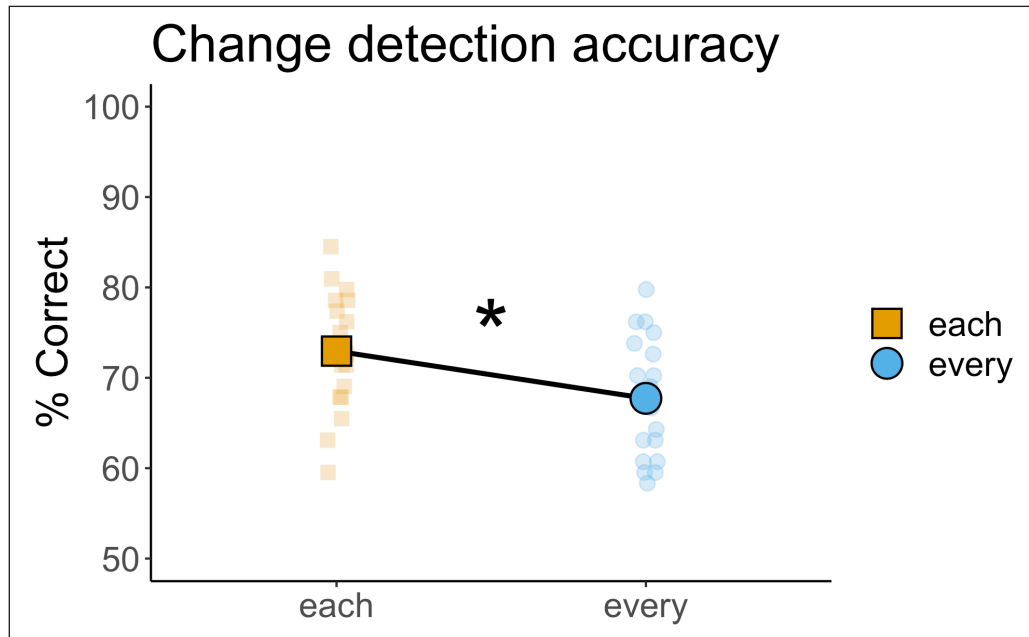
Figure 2: Experiment 1 performance. Participants were asked to evaluate the sentence *one circle changed its color* immediately after having evaluated a sentence like *each circle is green* (orange squares) or a sentence like *every circle is green* (blue circles). Points reflect performance on the change detection portion of the task. Translucent points represent each individual participant's accuracy. Significance star reflects p < .05.

Though participants were better at correctly rejecting the change detection statement than at recognizing a change when there was one, the *each* advantage did not interact with the type of follow-up question (trials on which there was a change versus trials on which there was no change). In particular, the interaction of quantifier and question type was not a significant predictor of accuracy in a binomial mixed effects regression model ($\beta = .05$ [95% CI –.15 to .25], $z = 0.25$, $p = .80$), fit using the lmerTest package for R (Kuznetsova et al., 2017). Moreover, the model including this interaction term did not result in a significantly better fit compared to a model that only included main effects of quantifier and question type ($\chi^2(1) = 0.06$, p = .803).

2.4 Discussion

As predicted, participants who initially evaluated *each*-statements performed better, on average, than those who initially evaluated *every*-statements. As the only difference between the two conditions was the quantifier used in the initial verification task, we can feel confident in

holding the meaning of those quantifiers responsible for this difference in performance. Indeed, this pattern of performance is well-explained if the expression *each circle* has a meaning that encouraged participants to represent individual circles whereas *every circle* has a meaning that encouraged participants to group the circles as an ensemble collection. This result is striking, given that there were only three circles on the screen. Three is well within adults' working memory capacity, so they could have chosen to encode the individual colors. After all, they knew they would have to answer a change detection question on every trial, so they would do well to take time to encode each circle's color. Instead, it seems that the linguistic framing led participants in the *every* condition to adopt a strategy that is sub-optimal with respect to this task.

That said, the effect size in Experiment 1 appears to be relatively small, likely owing to individual differences in change detection acuity and the between-subjects nature of the design. In Experiment 2, we aimed to find a stronger signal of the effect, by measuring participants' sensitivity to change detection and adjusting the task difficulty accordingly.

3. Experiment 2: Staircased difficulty

3.1 Participants

37 participants were recruited on Amazon Mechanical Turk. All gave informed consent and passed an English-screener prior to participating in the actual task. One participant was removed from further analysis for having response times longer than three standard deviations above the mean. The stimuli in Experiment 2 changed as a function of performance (as described below), so no participants were removed based on percent "correct" on the sentence verification portion of the task (though the results reported below are unchanged if the 55% exclusion criterion from Experiment 1 is applied). This left 36 participants.

3.2 Stimuli and procedure

Both the materials and the procedure were identical to those in Experiment 1, with one exception: difficulty changed as a function of participants' performance. Instead of randomly sampling hues for the change detection question from a Gaussian distribution centered over the original hue with a standard deviation of 17, the standard deviation of this distribution changed from trial to trial. At the start of the experiment it was set to 20 (a larger distribution (i.e., easier) than in Experiment 1). Any time the participant correctly detected a change, this value was decremented by one, making subsequent trials harder (i.e., the new hue was more likely to be close to the original hue). Any time the participant missed a change, this value was incremented by one, making subsequent trials easier (i.e., the new hue was more likely to be distant from the original hue). No change in trial difficulty occurred when participants correctly rejected the change question or falsely reported a change on a trial in which no change occurred. That is, the standard deviation remained the same on trials without a color change.

3.3 Results

Given the empirically determined color categories used to generate trials, participants correctly evaluated 65.9% of the *each*-statements and 65.4% of the *every*-statements. As noted above, this low "accuracy" was expected, given that hues on the border between color categories were used. On average, participants viewed the displays for 1739 milliseconds in the *every* condition and 2262 milliseconds in the *each* condition. As in Experiment 1, this difference between reaction times was not significant ($t_{34.33} = 1.73, p = .093$).

On the change detection portion of the task, difficulty varied based on performance, so participants achieved similar accuracies: 71.8% in the *each* condition and 71.2% in the *every* condition (this difference was not significant: $t_{32.00} = 0.28, p = .782$). But participants in the *every* condition required easier trials to achieve this level of accuracy. As seen in **Figure 3**, participants who initially evaluated *each*-sentences had a smaller standard deviation throughout the experiment – reflecting harder trials – than participants who initially evaluated *every*-sentences.
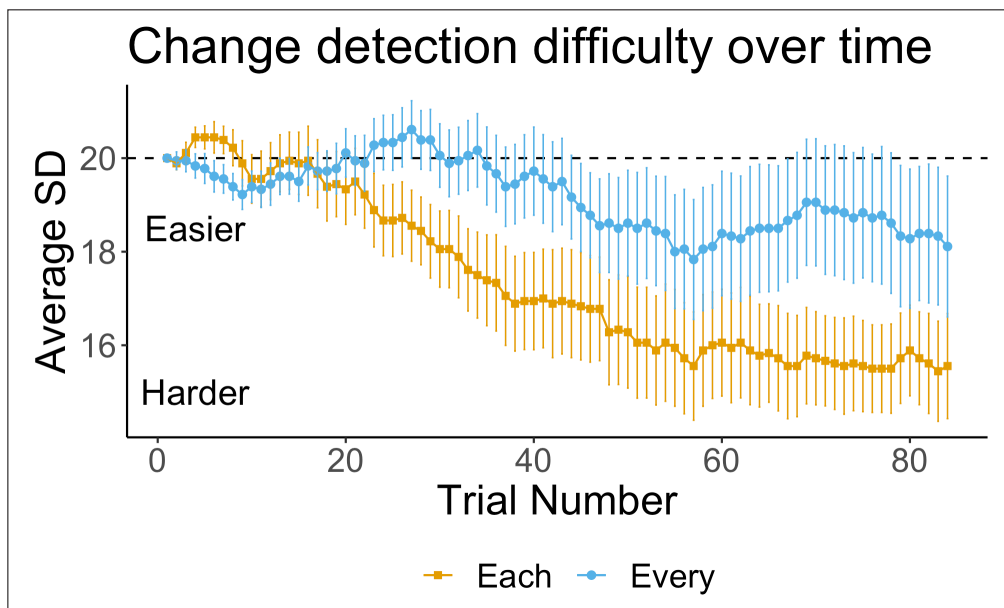
Figure 3: Experiment 2 trial difficulty. The standard deviation of the distribution from which the new color was chosen on each trial in the *each* condition (orange squares) and the *every* condition (blue circles). A larger standard deviation corresponds to an easier trial, on average, since the new hue is more likely to be farther from the original hue. Dotted line represents the starting difficulty: a standard deviation of 20. Error bars represent standard error.

To verify this statistically, a mixed-effects regression model was fit using the lmerTest package for R. Difficulty level (i.e., standard deviation) on a particular trial was predicted by trial number, quantifier (*each* = –.5; *every* = .5), and their interaction. Random intercepts for participants were included. We tested for significance of the interaction between trial number and quantifier by conducting a likelihood ratio test on the Chi-square values from model comparison (comparing the aforementioned model with the interaction term against a model that only included trial number as a predictor). This confirmed that the quantifier used significantly affected the difficulty level participants achieved ($\chi^2(1)$ = 136.94, p < .001; interaction of trial number and quantifier: β = .045 [95% CI .041 to .049], t = 11.7, p < .001).

As in Experiment 1, the type of change detection trial (one in which there was a change, as opposed to one in which there was no change) seemed to make no difference to the *each* advantage. When an interaction term is included in the model, it fails to be significant (interaction of quantifier and question type: β = –.19 [95% CI –.379 to –.002], t = 1.01, p = .311).

3.4 Discussion

The results of Experiment 2 corroborate the findings of Experiment 1 and demonstrate a different signal of the effect. As in Experiment 1, participants who first evaluated an *each*-sentence performed better in the subsequent change detection task than those who first evaluated a truth-conditionally equivalent *every*-sentence. Because the dependent measure in Experiment 2 was difficulty, both groups of participants achieved similar accuracies. But those in the *each* condition were able to do so even given significantly harder trials (i.e., ones in which the changes were fewer hues apart).

4. General discussion

We started with the question of how deep linguistic framing effects run. Closely related expressions like *chase* and *flee* are liable to inspire different thoughts in the minds of speakers, even when used to describe the same state of the world (*the dog chased the cat* and *the cat fled the dog*). But what about logically equivalent expressions like *each* and *every*? We find evidence that the difference in meaning between these quantifiers has predictable downstream consequences for how people represent the things quantified over. Participants were more accurate at detecting a change in a particular circle's hue when the relevant circles were introduced with *each,* compared to when they were introduced with *every* (Experiment 1). They were likewise able to reliably answer harder change detection questions following *each*-sentences than following *every*-sentences (Experiment 2).

These two experiments, in conjunction with past work, support the conclusion that sentences like *each circle is green* invite (but don't require) participants to represent the circles as individual object-files. Consequently, those circles' properties are encoded when participants view the corresponding scene, and participants can reliably detect when one of these properties (hue) changes. By contrast, these results (along with prior work) support the idea that minimally different sentences like *every circle is green* encourage participants to represent the circles as an ensemble. And in doing so, participants fail to encode those circles' individual properties as reliably. So those who evaluate an *every*-sentence are comparatively worse at detecting the color change of a particular circle (these results would predictably flip if the same displays were shown, but the follow-up question instead probed the average hue).

Importantly, only the quantifier differs between conditions in these experiments. And this difference in quantifier makes no difference to the truth or falsity of the original sentence. Given that the context restricts the domain of quantification to just the circles on the screen, anytime it's true that *each circle is green*, it's likewise true that *every circle is green* (and likewise true that *every single one of the circles is green*). Nonetheless, the choice of quantifier affects how participants approach the scene, encouraging them to represent the circles as independent individuals in the *each* case and as a collection in the *every* case. And whereas previous findings of enhanced performance on a follow-up memory question for *every* over *each* might have been explainable in terms of increased processing costs for *each* (perhaps due to its relative infrequency), the current result of *each* leading to better performance than *every* militates against this low-level concern.

A reviewer raises the point that if participants initially encoded the circles as an ensemble in the *every* case, they could have succeeded at the change detection question by comparing the change in average hue between both screens. This is, in principle, possible, and we suspect participants could be invited to deploy this sort of strategy if both displays of circles were described in ways that promoted ensemble representation (e.g., "every circle is green … did every circle stay the same color?"). But, as it is, the second display was introduced in a way that invited the circles to be individuated (another case of linguistic framing altering visual object construal). And it is this mismatch – between a push toward ensemble representation on the initial display and a push toward object individuation on the subsequent display – that we believe accounts for the relatively poor performance given *every*.

To explain this result and suggest a unified explanation of the previously observed contrasts, we proposed that *each* and *every* are understood along the lines of (2b) and (2c). In particular, *each* has a first-order (i.e., purely individualistic) meaning, whereas *every* expresses the same content – distributive universal quantification – but with a second-order twist: a call for grouping the things that satisfy the noun phrase with which the quantifier combines. This difference in representational format, on our view, results in *each* inviting (but not entailing) representing the domain as a series of independent individuals, whereas *every* invites (but does not entail) ensemble representation. As

noted in Section 1, we take this non-deterministic link between linguistic meanings and non-linguistic cognition to be a feature of the present proposal, as it captures the logical equivalence of these two quantifiers, while also explaining the non-categorical patterns of use that differentiate them.

To be sure, the results presented here are between-subjects comparisons, so they cannot license the strong conclusion that every speaker of English understands *each* and *every* along the lines we propose.[5] However, the alternative that only most speakers understand *each* and *every* in the proposed way does not strike us as a particularly parsimonious position. Such a view would call for an explanation of why some people come to acquire different meanings for what we think of as the same word. If it were true that some people pair *each* with a meaning that gets used to build thoughts like (2b), while others pair it with a meaning that gets used to build thoughts like (2c), one would want to know what differences in their input led them to such a conclusion. In this context, data pertaining to the acquisition of *each* and *every* will be useful to consider (see, e.g., Knowlton & Gomes, 2022; Knowlton & Lidz, 2021; Rasin & Aravind, 2021).

That said, the intuition that *each* is somehow more individualistic than *every* is an old one. And while it is generally put to the side – even in explicit theoretical proposals about quantifier meanings (e.g., Champollion, 2017; Winter, 2002) – there are theories which have taken aim at some of the differences between *each* and *every* discussed in 1.1. We briefly turn to two such proposals and ask whether they could accommodate the present results as well.

4.1 Do alternative accounts make identical predictions?

One prominent example of a view that takes the differences between *each* and *every* seriously comes from Beghelli and Stowell (1997). They aim to account for some of the differences discussed in 1.1 without supposing the two quantifiers differ in meaning, but by proposing distinct syntactic features be appended to the lexical items. This allows theorists to retain the standard treatment of quantifiers (e.g., Barwise & Cooper, 1981) at the expense of adding syntactic complexity. What follows is a very general sketch of their proposal, abstracting away from much of its rich detail.

The gist of Beghelli and Stowell's view is that *each* is marked with a "strong distributivity" feature, causing it to move higher in the syntactic tree to associate with a distributivity operator, which is responsible for enforcing distributive readings of sentences. That is, the distributivity operator ensures that predicates apply to individuals. Since *each* always undergoes movement to associate with this operator, *each* is always distributive (cp. the related idea that as opposed to associating with an otherwise unpronounced distributivity operator, *each* is the pronunciation of that operator; LaTerza, 2014). On the other hand, *every* has a "weak distributivity" feature,

---

[5] Moreover, given the non-deterministic nature of the proposed link between linguistic meanings and extralinguistic cognition – the bias to represent object-files or ensembles is not a hard-and-fast rule – data from a within-subjects comparison would also likely fail to underwrite or undermine that strong conclusion.

meaning that this movement and subsequent association with the distributivity operator usually occurs, but can be prevented. This is what is argued to happen in cases where *every* resists pair-list readings, such as (6) from 1.1: "Which book did you loan to every student?".

The distributivity operator also happens to be located higher in the syntactic tree than the generic operator, which is responsible for giving sentences generic meanings (e.g., *cakes need sugar*). So, for Beghelli and Stowell, *every* is sometimes amenable to such interpretations (e.g., *every cake needs sugar* sounds like something you might find in a recipe book) because *every NP* sometimes scopes below the generic operator. On the other hand, *each* always takes scope above this operator and thus resists generic interpretations (*each cake needs sugar* sounds like a claim about particular cakes).

Could this cartographic account capture the results we report here? That is, could it be that *each* mandatorily associating with the distributivity operator is responsible for our result that *each* triggers object-file representations? We think not. At least in our experimental context, there is no non-distributive way for every circle to be green. Whatever one thinks about *each*, *every*, and the distributivity operator, the sentence *every circle is green* is no less distributive than *each circle is green*. Whether this is because *every* does associate with the distributivity operator in that sentence or because the meaning of the distributivity operator is implicit in the meaning of the predicate *be green*, it seems clear that both sentences are equally distributive. But if both sentences are equally distributive, then any distributivity provided by *each* associating with the distributivity operator in the case of *each circle is green* would be completely redundant in terms of its behavioral import. That is, if association with distributivity gives rise to object-file representation, then both conditions should equally give rise to object-file representation, because distributivity is present in both.

Moving to a more lexically-oriented view, Tunstall (1998) also takes seriously the differences between *each* and *every*, proposing that they share a common core but differ with respect to the conditions they place on the events they describe (see also Brasoveanu & Dotlačil, 2015). According to Tunstall's *event differentiation condition*, *each* requires each individual in the denotation of its internal argument to be associated with its own event, in some sense. On this view, *Kermit lifted each box* must describe a situation in which Kermit lifted each box independently of all others. It cannot truthfully describe a situation in which Kermit lifted box_1, and then lifted box_2 and box_3 simultaneously. In contrast, *every* is said to have the weaker requirement that there be at least two distinct events (dubbed the *event distributivity condition*). So *Kermit lifted every box* could describe either of the aforementioned situations, but not one in which Kermit lifted all three boxes at once. We do not feel that these judgements reflect a difference in *truth-conditions* (though they might constitute further evidence of a representational distinction). But in any case, assuming some version of Tunstall's proposal is on the right track, it raises the question: could such an event differentiation condition capture the above results?

In this context, it is more difficult to imagine the linking hypothesis that connects observed performance in an experiment to a particular claim about how to explain that performance. The linking hypothesis adopted throughout this paper comes from Lidz et al. (2011): the verification procedures employed in understanding a declarative sentence are biased toward algorithms that directly compute the relations and operations expressed by the semantic representation of that sentence. In other words, meanings, like all mental representations, have particular formats (i.e., they express particular relations and operations). Those formats highlight the applicability of certain computational procedures and background others.

In this sense, the linking hypothesis is reminiscent of Marr (1982)'s discussion of algorithmic-level distinctions. For example, the numerical content THIRTY-SEVEN could be represented in a base-10 system, as "37", or in a base-2 system, as "100101". Whereas the former makes decomposition into powers of 10 explicit, the latter makes decomposition into powers of 2 explicit. In the same vein, *each* and *every* express universal content that might be represented in first-order terms, as in (2b), or second-order terms, as in (2c). Whereas the first-order (2b) makes each individual circle explicit, the second-order (2c) makes the group of circles explicit.

The connection between the proposed representations in (2) and the predicted performance follows from the linking hypothesis discussed above. Namely, understanding *each circle is green* amounts to building a representation like (2b). This representation highlights individual circles. Object-files are representations of individuals as such. So when participants want to evaluate this representation for truth, they are biased (but not required) to do so with a strategy that involves representing individual circles as object-files. This relatively straightforward linking hypothesis has received empirical support from case studies not involving universal quantifiers (e.g., Knowlton, Hunter, et al., 2021; Lidz et al., 2011; Odic et al., 2018; Pietroski et al., 2009; Tomaszewicz, 2011, 2013; Tomaszewicz-Özakın, 2021).

Returning to Tunstall's event differentiation condition from above, we might ask why a representation that treats events as partially distinct versus fully distinct would be expected to eventuate in ensemble versus object-file representations of the domain. We see no obvious answer to that question. So while our results may not be inconsistent with Tunstall's account – *each* and *every* may well impose different conditions on the events they describe – they are certainly not predicted by it.

4.2 Conclusion

The results reported here suggest that linguistic framing interacts in theoretically interesting ways with language-independent representational systems. In doing so, they tell against the standard view that what speakers know when they know the meaning of a lexical item like *each* or *every* is the contribution that it makes to the truth-conditions of sentences (e.g., that greenness

applies to circles universally as opposed to existentially). On such views, how speakers ultimately mentally represent that contribution is taken to be of less importance, either because it is thought to vary from person to person or because speakers are taken to represent an equivalence class of specifications of the truth-condition in question. But our results suggest that what serves as the meaning of an expression like *each* or *every* is a representation with a particular format. The meanings of these quantifiers both provide precise instructions to assemble thoughts with particular formal structures, which in turn have downstream consequences on thought (e.g., triggering object-file or ensemble representation; affecting linguistic use in interesting ways).

To be clear, no one is likely to doubt that different expressions can be used to describe the same referent (or situation), yet inspire distinct thoughts. Pairs of expressions like *woodchuck* and *groundhog* – or *Hesperus* and *Phosphorus*, or *creatures with a heart* and *creatures with a kidney* – can have different associations in the minds of speakers who use those expressions. Theorists disagree about whether such expressions share a meaning, in part because theorists disagree about what meanings are. But it seems obvious that thoughts tokened in response to encountering "co-extensive" expressions might differ, if only because speakers might link such expressions with different histories of use.

However, the framing effect we observe here is not a mere symptom of *each* and *every* having different associations. It arises from a formal distinction in the expressions' meanings and a principled relationship between those psycho-logical forms and non-linguistic representational systems. We can make sense of what gets encoded during sentence evaluation (individual details versus group summary statistics) by appealing to the distinction between first-order and second-order representations and the idea that these logically equivalent representations often trigger distinct cognitive systems. Namely, the first-order meaning of *each* naturally interfaces with (but does not depend on) the system for representing object-files. On the other hand, the second-order meaning of *every* interfaces with (but does not depend on) the system for representing ensembles. It is for this reason that we take the linguistic framing effect discussed here to be a relatively "deep" one.

Vendler (1962), in his original study of universal quantifiers, noted that any differences between *each* and *every* "are much too fine to be located by merely comparing truth-values." He continued: "In order to spot them we have to summon our best feeling for English idioms, and without disdaining help from other quarters…" (p.148). We hope to have shown that evidence from visual object construal constitutes one such quarter. Moreover, this case study on universal quantifiers supports the idea that details about how participants verify sentences in carefully controlled settings can be useful for drawing inferences about the form of the mental representations that serve as expressions' meanings.

Data accessibility statement

Stimuli, data, and analysis scripts are available here: https://osf.io/f78er/?view_only=54f0fbd526074e2ab2239096c10959c0

Ethics and consent

The experiments reported here were approved by The University of Maryland's IRB.

Acknowledgements

Competing interests

The authors have no competing interests to declare.

Authors' contributions

T.K., J.H., P.P., and J.L. conceptualized the project and designed the experiments. T.K. collected the data and analyzed it with input from J.H., P.P., and J.L. T.K. wrote the manuscript with input from J.H., P.P., and J.L.

References

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences, 15*(3), 122–131. DOI: https://doi.org/10.1016/j.tics.2011.01.003

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science, 19*(4), 392–398. DOI: https://doi.org/10.1111/j.1467-9280.2008.02098.x

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*(2), 157–162. DOI: https://doi.org/10.1111/1467-9280.00327

Bae, G.-Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General, 144*(4), 744–763. DOI: https://doi.org/10.1037/xge0000076

Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. In *Linguistics and Philosophy*, *4*, 159–219. DOI: https://doi.org/10.1007/BF00350139

Beghelli, F. (1997). The syntax of distributivity and pair-list readings. In A. Szabolcsi (Ed.) *Ways of scope taking* pp. 349–408. Springer, Dordrecht. DOI: https://doi.org/10.1007/978-94-011-5814-5_10

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. Springer, Dordrecht. DOI: https://doi.org/10.1007/978-94-011-5814-5_3

Brasoveanu, A., & Dotlačil, J. (2015). Strategies for scope taking. *Natural Language Semantics 23*(1), 1–19. DOI: https://doi.org/10.1007/s11050-014-9109-1

Carey, S. (2009). *The origin of concepts*. Oxford Series in Cognitive Development. Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780195367638.001.0001

Champollion, L. (2017). *Parts of a whole: Distributivity as a bridge between aspect and measurement*. Oxford University Press. DOI: https://doi.org/10.1093/oso/9780198755128.003.0009

Chomsky, N. (1964). *Current issues in linguistic theory*. De Gruyter Mouton. DOI: https://doi.org/10.1515/9783110867565

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404. DOI: https://doi.org/10.1016/S0042-6989(02)00596-5

Clark, R., & Grossman, M. (2007). Number sense and quantifier interpretation. *Topoi*, *26*, 51–62. Springer. DOI: https://doi.org/10.1007/s11245-006-9008-2

Duchon, D., Dunegan, K. J., & Barton, S. L. (1989). Framing the problem and making decisions: The facts are not enough. *IEEE Transactions on Engineering Management*, *36*(1), 25–27. DOI: https://doi.org/10.1109/17.19979

Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Developmental Science*, *6*(5), 568–584. DOI: https://doi.org/10.1111/1467-7687.00313

Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition*, *97*(3), 295–313. DOI: https://doi.org/10.1016/j.cognition.2004.09.010

Feigenson, L., Dehaene, S., & Spelke, E. S. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*, 307–314. DOI: https://doi.org/10.1016/j.cogpsych.2016.04.002

Feiman, R., & Snedeker, J. (2016). The logic in language: How all quantifiers are alike, but each quantifier is different. *Cognitive Psychology*, *87*, 29–52. DOI: https://doi.org/10.1016/j.cogpsych.2016.04.002

Geurts, B. (2013). Alternatives in framing and decision making. *Mind & Language*, *28*(1), 1–19. DOI: https://doi.org/10.1111/mila.12005

Gil, D. (1992). Scopal quantifiers: Some universals of lexical effability. In M. Kefer & J. van der Auwera (Eds.), *Meaning and grammar: Cross-linguistic perspectives* pp. 303–345. Mouton de Gruyter.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55. DOI: https://doi.org/10.1207/s15327817la0101_2

Green, E. J., & Quilty-Dunn, J. (2021). What is an object file? *The British Journal for the Philosophy of Science*, *72*(3), 665-699. DOI: https://doi.org/10.1093/bjps/axx055

Hornstein, N. (1984). *Logic as grammar*. MIT Press. DOI: https://doi.org/10.7551/mitpress/4287.001.0001

Haberman, J., & Whitney, D. (2011). Efficient summary statistical representation when change localization fails. *Psychonomic Bulletin & Review*, *18*(5), 855–859. DOI: https://doi.org/10.3758/s13423-011-0125-6

Haberman, J., & Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. In J. Wolfe & L. Robertson (Eds.) *From perception to consciousness: Searching with Anne Treisman* pp. 339-349. Oxford University Press. DOI: https://doi.org/10.1093/acprof:osobl/9780199734337.003.0030

Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences, 106*(25), 10382–10385. DOI: https://doi.org/10.1073/pnas.0812142106

Jackendoff, R. S. (1983). *Semantics and cognition*. MIT Press.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity in varieties of attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* pp. 29–61. Academic Press.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*(2), 175–219. DOI: https://doi.org/10.1016/0010-0285(92)90007-O

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica, 47*, 263–291. DOI: https://doi.org/10.2307/1914185

Kanjlia, S., Feigenson, L., & Bedny, M. (2021). Neural basis of approximate number in congenital blindness. *Cortex, 142*, 342-356. DOI: https://doi.org/10.1016/j.cortex.2021.06.004

Knowlton, T. (2021). *The psycho-logic of universal quantifiers* [Doctoral dissertation]. University of Maryland. DOI: https://doi.org/10.13016/fdr8-3qqh

Knowlton, T., & Gomes, V. (2022). Linguistic and non-linguistic cues to acquiring the strong distributivity of "each". *Proceedings of the Linguistic Society of America*, *7*(1). DOI: https://doi.org/10.3765/plsa.v7i1.5236

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. 1*, 134–144. DOI: 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 *Proceedings of the Boston University Conference on Language Development, 45*, 399–412. http://www.lingref.com/bucld/45/BUCLD45-31.pdf

Knowlton, T., Pietroski, P., Halberda, J., & Lidz, J. (2022). The mental representation of universal quantifiers. *Linguistics and Philosophy, 45*, 911–941. DOI: https://doi.org/10.1007/s10988-021-09337-8

Knowlton, T., Pietroski, P., Williams, A., Halberda, J., & Lidz, J. (2021). Determiners are 'conservative' because their meanings are not relations: evidence from verification. *Semantics and Linguistic Theory, 30,* 206–226. DOI: https://doi.org/10.3765/salt.v30i0.4815

Knowlton, T., Trueswell, J., & Papafragou, A. (2023). Keeping quantifier meaning in mind: Connecting semantics, cognition, and pragmatics. *Cognitive Psychology, 144,* 101584. DOI: https://doi.org/10.1016/j.cogpsych.2023.101584

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. DOI: https://doi.org/10.18637/jss.v082.i13

Landman, F. (2003). Predicate-argument mismatches and the adjectival theory of indefinites. In M. Coene & Y. D'hulst (Eds.), *From NP to DP, 1,* 211–237. J. Benjamins Publishing Company. DOI: https://doi.org/10.1075/la.55.10lan

LaTerza, C. (2014). *Distributivity and plural anaphora* [Doctoral dissertation]. University of Maryland. http://hdl.handle.net/1903/15826

Lidz, J., Pietroski, P., Halberda, J., & Hunter, T. (2011). Interface transparency and the psychosemantics of most. *Natural Language Semantics, 19*(3), 227–256. DOI: https://doi.org/10.1007/s11050-010-9062-6

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

May, R. (1985). *Logical form: Its structure and derivation.* MIT Press.

McMillan, C., Clark, R., Moore, P., Devita, C., & Grossman, M. (2005) Neural basis for generalized quantifiers comprehension. *Neuropsychologia, 43,* 1729–1737. DOI: https://doi.org/10.1016/j.neuropsychologia.2005.02.012

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). DOI: https://doi.org/10.5334/gjgl.409

Olm, C. A., McMillan, C. T., Spotorno, N., Clark, R., & Grossman, M. (2014). The relative contributions of frontal and parietal cortex for generalized quantifier comprehension. *Frontiers in Human Neuroscience, 8.* DOI: https://doi.org/10.3389/fnhum.2014.00610

Piazza, E. A., Sweeny, T. D., Wessel, D., Silver, M. A., & Whitney, D. (2013). Humans use summary statistics to perceive auditory sequences. *Psychological Science, 24*(8), 1389-1397. DOI: https://doi.org/10.1177/0956797612473759

Pietroski, P. M. (2018). *Conjoining meanings: Semantics without truth values.* Oxford University Press. DOI: https://doi.org/10.1093/oso/9780198812722.001.0001

Pietroski, P., Lidz, J., Hunter, T., & Halberda, J. (2009). The meaning of 'most': Semantics, numerosity and psychology. *Mind & Language, 24*(5), 554–585. DOI: https://doi.org/10.1111/j.1468-0017.2009.01374.x

Plaisier, M. A., Tiest, W. M. B., & Kappers, A. M. (2009). One, two, three, many – Subitizing in active touch. *Acta psychologica, 131*(2), 163–170. DOI: https://doi.org/10.1016/j.actpsy.2009.04.003

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition, 80*(1-2), 127-158. DOI: https://doi.org/10.1016/S0010-0277(00)00156-6

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3*(3), 179–197. DOI: https://doi.org/10.1163/156856888X00122

Rasin, E., & Aravind, A. (2021). The nature of the semantic stimulus: The acquisition of *every* as a case study. *Natural Language Semantics, 29,* 339–375. DOI: https://doi.org/10.1007/s11050-020-09168-6

Riggs, K. J., Ferrand, L., Lancelin, D., Fryziel, L., Dumur, G., & Simpson, A. (2006). Subitizing in tactile perception. *Psychological Science, 17*(4), 271–272. DOI: https://doi.org/10.1111/j.1467-9280.2006.01696.x

Soames, S. (1987). *Logic as grammar* by Norbert Hornstein. *The Journal of Philosophy, 84*(8), 447–455. DOI: https://doi.org/10.5840/jphil198784891

Surányi, L. B. (2003). *Multiple operator movements in Hungarian* [Doctoral dissertation]. Utrecht University. https://dspace.library.uu.nl/handle/1874/623

Sweeny, T. D., Wurnitsch, N., Gopnik, A., & Whitney, D. (2015). Ensemble perception of size in 4–5-year-old children. *Developmental Science, 18*(4), 556–568. DOI: https://doi.org/10.1111/desc.12239

Szabolcsi, A. (2010). *Quantification.* Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511781681

Szymanik, J. (2016). *Quantifiers and cognition: Logical and computational perspectives.* Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-28749-2

Szymanik, J., & Zajenkowski, M. (2010). Comprehension of simple quantifiers: Empirical evaluation of a computational model. *Cognitive Science, 34*(3), 521–532. DOI: https://doi.org/10.1111/j.1551-6709.2009.01078.x

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

Tomaszewicz, B. (2013). Linguistic and visual cognition: Verifying proportional and superlative *most* in Bulgarian and Polish. *Journal of Logic, Language and Information, 22*(3), 335–356. DOI: https://doi.org/10.1007/s10849-013-9176-6

Tomaszewicz-Özakın, B. (2021). *Some, most, all* in a visual world study. *Formal Approaches to Number in Slavic and Beyond, 5,* 399.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211,* 453–458. DOI: https://doi.org/10.1126/science.7455683

Tunstall, S. (1998). *The interpretation of quantifiers: Semantics and processing* [Doctoral dissertation]. University of Massachusetts Amherst. https://scholarworks.umass.edu/dissertations/AAI9909228

van Benthem, J. (1986). *Essays in logical semantics.* Dordrecht: D. Reidel Publishing Company. DOI: https://doi.org/10.1007/978-94-009-4540-1

Vendler, Z. (1962). Each and every, any and all. *Mind*, 71(282), 145–160. DOI: https://doi.org/10.1093/mind/LXXI.282.145

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27*(1), 92–114. DOI: https://doi.org/10.1037/0096-1523.27.1.92

Ward, E. J., Bear, A., & Scholl, B. J. (2016). Can you perceive ensembles without perceiving individuals?: The role of statistical perception in determining whether awareness overflows access. *Cognition, 152,* 78–86. DOI: https://doi.org/10.1016/j.cognition.2016.01.010

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology, 69,* 105–129. DOI: https://doi.org/10.1146/annurev-psych-010416-044232

Winter, Y. (2002). Atoms and sets: A characterization of semantic number. *Linguistic Inquiry, 33*(3), 493–505. DOI: https://doi.org/10.1162/002438902760168581

Wood, J. N., & Spelke, E. S. (2005). Infants' enumeration of actions: Numerical discrimination and its signature limits. *Developmental Science, 8*(2), 173–181. DOI: https://doi.org/10.1111/j.1467-7687.2005.00404.x

Xu, F., & Carey, S. (1996). Infants metaphysics: The case of numerical identity. *Cognitive Psychology, 30*(2), 111–153. DOI: https://doi.org/10.1006/cogp.1996.0005

Xu, Y., & Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends in cognitive sciences, 13*(4), 167–174. DOI: https://doi.org/10.1016/j.tics.2009.01.008

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*(1), B1–B11. DOI: https://doi.org/10.1016/S0010-0277(99)00066-9