ABSTRACT

Title of Dissertation: THE PSYCHO-LOGIC
OF UNIVERSAL QUANTIFIERS

Tyler Knowlton
Doctor of Philosophy, 2021

Dissertation Directed by: Professor Jeffrey Lidz
Department of Linguistics

Professor Emeritus Paul Pietroski
Department of Linguistics

A universally quantified sentence like *every frog is green* is standardly thought
to express a two-place second-order relation (e.g., *the set of frogs is a subset of the
set of green things*). This dissertation argues that as a psychological hypothesis
about how speakers mentally represent universal quantifiers, this view is wrong in
two respects. First, *each*, *every*, and *all* are not represented as two-place relations,
but as one-place descriptions of how a predicate applies to a restricted domain (e.g.,
relative to the frogs, everything is green). Second, while *every* and *all* are represented
in a second-order way that implicates a group, *each* is represented in a completely
first-order way that does not involve grouping the satisfiers of a predicate together
(e.g., *relative to individual frogs, each one is green*).

These "psycho-logical" distinctions have consequences for how participants
evaluate sentences like *every circle is green* in controlled settings. In particular,
participants represent the extension of the determiner's internal argument (the cir-

cles), but not the extension of its external argument (the green things). Moreover, the cognitive system they use to represent the internal argument differs depending on the determiner: Given *every* or *all*, participants show signatures of forming ensemble representations, but given *each*, they represent individual object-files.

In addition to psychosemantic evidence, the proposed representations provide explanations for at least two semantic phenomena. The first is the "conservativity" universal: All determiners allow for duplicating their first argument in their second argument without a change in informational significance (e.g., *every fish swims* has the same truth-conditions as *every fish is a fish that swims*). This is a puzzling generalization if determiners express two-place relations, but it is a logical consequence if they are devices for forming one-place restricted quantifiers.

The second is that *every*, but not *each*, naturally invites certain kinds of generic interpretations (e.g., *gravity acts on every/#each object*). This asymmetry can potentially be explained by details of the interfacing cognitive systems (ensemble and object-file representations). And given that the difference leads to lower-level concomitants in child-ambient speech (as revealed by a corpus investigation), children may be able to leverage it to acquire *every*'s second-order meaning.

This case study on the universal quantifiers suggests that knowing the meaning of a word like *every* consists not just in understanding the informational contribution that it makes, but in representing that contribution in a particular format. And much like phonological representations provide instructions to the motor planning system, it supports the idea that meaning representations provide (sometimes surprisingly precise) instructions to conceptual systems.

THE PSYCHO-LOGIC OF UNIVERSAL QUANTIFIERS

by

Tyler Knowlton

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Advisory Committee:
Professor Jeffrey Lidz, Chair
Professor Emeritus Paul Pietroski, Co-Chair
Professor Alexander Williams
Professor Valentine Hacquard
Professor Justin Halberda
Professor Yi Ting Huang, Dean's Representative

Preface

The work reported in this dissertation is collaborative. Chapter 2 reports on joint work with Jeffrey Lidz, Paul Pietroski, Justin Halberda, and Alexander Williams. Portions of it were published in Knowlton et al. (2021b). Chapter 3 reports on joint work with Jeffrey Lidz, Paul Pietroski, and Justin Halberda. Portions of it were published in Knowlton et al. (accepted). Chapter 4 reports on joint work with Jeffrey Lidz. Portions of it were published in Knowlton and Lidz (2021). This work was supported by the National Science Foundation (Doctoral Dissertation Research Improvement grant #BCS-2017525 and NRT award #DGE-1449815) and by the James S. McDonnell foundation (Collaborative Activity Award on *The Nature and Origins of the Human Capacity for Abstract Combinatorial Thought*).

Acknowledgments

There are so many people to thank. To begin with, I've been lucky to have mentors who, from the start, treated me as a collaborator and a friend. I'm immensely grateful to them, in a way that I won't be able to do justice to here.

Thanks to Jeff Lidz for teaching me that slow is fast, and for also teaching me how to slow down and think. When I first met with Jeff, I felt the need to fill any gap in the conversation with whatever half-formed thoughts came to mind. But by the end I learned the important skill of being comfortable with thinking together in silence. I'm told that some mentors take a "you generate, I filter" approach; not Jeff. Jeff never rejected a bad idea that I brought him. Instead, he figured out how to improve it, refine it, or salvage some part of it. Even with all his guidance, he still made me feel that I had ownership over anything I was working on. And importantly, he made sure I was always having fun while working on it.

Thanks to Paul Pietroski for the profound influence he has had on my thinking about language, meaning, and science in general. This could have gone without saying, given how prominently his influence is felt throughout this dissertation. Whether in Maryland, New Brunswick, or New Mexico, Paul always made time to meet with me, help me think through arguments, and remind me about the bigger picture. He was patient with me when I asked him to explain, for the umpteenth time, Frege's view or Boolos' view or his view. He also gives fantastic advice about writing, balancing work and life, and mixing a drink.

Thanks to Justin Halberda for helping me discover this corner of cognitive science that I love. When I was an undergrad at Johns Hopkins, Justin hired me to manage his lab. But he made sure there was plenty of time for learning, exploring, and dipping my toes into different projects and collaborations. After a while, Justin invited me to tag along on his drives from Baltimore to College Park to meet with Jeff and Paul and talk about psychosemantics. The meetings were awesome. And I cherished those drives, where I learned about the language of thought, bathroom tiling patterns, and everything in between.

Thanks to Alexander Williams for showing me how to be a careful thinker. Alexander was always quick to lend me a sympathetic – but not uncritical – ear or eye. He'd offer excellent advice, and then he'd patiently offer it again after I inevitably failed to take it the first time. Also, his storytelling ability is second-to-none. This has been a great source of merriment and edification for me over the past few years.

Thanks to Valentine Hacquard for keeping me grounded and reminding me to pick my battles. Valentine has given me so many thoughtful comments, not only

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ANS	Approximate Number System
GQT	Generalized Quantifier Theory
ITT	Interface Transparency Thesis
LF	logical form

I	internal predicate/argument
E	external predicate/argument
DIST	distributivity feature
GEN	generic operator
e	entity type
v	event type
t	truth-value type
NP	noun phrase
DP	determiner phrase
QP	quantificational determiner phrase
VP	verb phrase

every	the expression "every"
EVERY	the "every" concept

iff	if and only if
s.t.	such that

Chapter 1:   Introduction

   This dissertation is about the representations that serve as the meanings of universal quantifiers, given a mentalistic view of meaning. In particular, the goal is to discriminate between representations that have the same truth-conditions. There are many plausible truth-conditionally equivalent specifications of a determiner like *every*. We might imagine that a sentence like *every frog is green* serves as an instruction to create the thought *the frogs are among the green things*, or the thought *the frogs are such that they are green*, or the thought *any thing that's a frog is such that it is green*. The focus in this dissertation will be on two "psycho-logical" distinctions that can be drawn between the possible specifications alluded to above: relational versus restricted quantification and first-order versus second-order quantification. The main question will be this: Which specifications most accurately describe speakers' mental representations?

   Posing this question assumes a mentalistic conception of linguistic meaning. In contrast, a popular tradition in semantics treats expressions as names for things in speakers' environments (e.g., Frege 1892; Davidson 1967b; Montague 1973; Lewis 1975; Heim and Kratzer 1998). On this view, a common noun like *frog* might be thought of as the name for a set (i.e., the set of frogs). Likewise, a quantificational

determiner like *every* might be thought of as the name for a relation between two sets (e.g., the set of frogs and the set of green things, in a sentence like (1a)). Theorists might choose to specify the relation expressed by *every* in any number of ways, including those in (1b) - (1e).

(1)   a. Every frog is green

        b. $\forall x(\mathbf{frog}(x))[\mathbf{green}(x)]$

            $\approx$ *each thing that is a frog is green*

        c. $\neg\exists x(\mathbf{frog}(x))[\neg\mathbf{green}(x)]$

            $\approx$ *no thing that is a frog is not green*

        d. $\{x : \mathbf{frog}(x)\} \subseteq \{x : \mathbf{green}(x)\}$

            $\approx$ *the set of frogs is a subset of the set of green things*

        e. $\{x : \mathbf{frog}(x)\} = \{x : \mathbf{frog}(x)\ \&\ \mathbf{green}(x)\}$

            $\approx$ *the set of frogs is identical to the set of green frogs*

These specifications are all logically equivalent, so they all do an equally fine job of capturing the truth-conditions of (1a). In that sense, each one of them can be thought of as a notational variant of the others, no different than writing the fancy $\mathcal{EVERY\ FROG\ IS\ GREEN}$ instead of the italicized *every frog is green*.

But while these specifications may be equivalent for purposes of describing truth-conditions, they are not equivalent as descriptions of how competent speakers understand a sentence like (1a). That is, (1b) - (1e) can be taken to represent distinct psychological hypotheses about how speakers mentally represent the sentence

in question. For example, (1b) could be represented by a mind incapable of representing negation; (1c) could not. Likewise, (1d) could be represented by a mind without the identity relation; (1e) could not. On the standard view, these differences in representational format are disregarded. The formalisms deployed by theorists are not meant to be related to the mental vocabulary that humans use to represent semantic properties of linguistic expressions (Dowty, 1979).

This is not, by itself, a criticism. Abstracting away from psychological considerations can be useful for certain purposes (e.g., exploring the compositional properties of meanings). But to the extent that natural language is a product of human minds and linguistics is a part of cognitive science, understanding the mental representations that serve as meanings in speakers' minds is a worthwhile project in its own right. This dissertation is an investigation into one corner of that project: the mental representation of the universal quantifiers *each*, *every*, and *all*.

The starting point will be a "psychologized" version of the standard view in semantics – Generalized Quantifier Theory – which holds that quantificational determiners are understood as two-place second-order relations (e.g., *the frogs are among the green things*). This dissertation argues that this view is wrong in two respects. First, universal quantifiers are not relational: *Every* and *all* are mentally represented as devices for creating one-place restricted quantifiers that are second-order (e.g., *the frogs are such that they are green*). And second, not all determiners are second-order: *Each* is mentally represented as a device for creating a one-place restricted quantifier that is first-order (e.g., *any individual that's a frog is such that it is green*). These proposed representations receive empirical support from psy-

cholinguistic experiments, offer explanations for linguistic phenomena, and suggest a novel path forward for studying language acquisition.

Zooming out, this investigation into the mental representation of universal quantifiers informs larger questions concerning the relationship between linguistic meanings and non-linguistic cognition. Meanings can be situated in mental grammar as the things that syntax connects with pronunciations (Chomsky, 1964). Pronunciations can be thought of as instructions to motor planning systems (Halle, 2003; Liberman and Mattingly, 1985; Poeppel et al., 2008). For example, the phonemic representation of *every frog is green* can be understood to be a collection of details about how to position the lips and tongue and how to regulate airflow. Meanings, on analogy, can be thought of as instructions to cognitive systems (Pietroski, 2018). For example, the semantic representation of *every frog is green* can be understood to be a collection of details about how to assemble a thought by combining certain concepts in a particular way.

Given this unabashedly mentalistic picture, we can ask questions like: What sorts of instructions do meanings provide to cognition? At what grain-size are these instructions shared by competent speakers? To what extent do the instructions supplied by the meaning constrain the contours of the resulting thought? This dissertation supports the idea that, at least in the context of "logical" vocabulary like quantificational determiners, meanings can offer surprisingly precise constraints on thought-building. And in doing so, it supplies grist for the mill of the mentalistic project.

1.1 Semantic claims as psychological hypotheses

As noted above, semantic claims often abstract away from details concerning mental representation. In particular, formal semantic theories focus on capturing the truth-conditional content of expressions (in a compositional way) and rarely make claims about the format in which that content is represented. For example, consider a common noun like *frog*. On standard treatments (e.g., Frege 1892; Davidson 1967b; Montague 1973; Lewis 1975; Heim and Kratzer 1998), *frog* is taken to express a function from entities to truth values, returning TRUE if the entity is a frog and FALSE otherwise. The extension of that function is the set of input/output pairs in (2a). Or, restricting our attention to just the cases where the function outputs TRUE, we can say that its extension is the set of frogs in (2b).

(2)   a. $\{..., \langle Kermit, \text{TRUE}\rangle, \langle Mr.Toad, \text{TRUE}\rangle, \langle Gritty^1, \text{FALSE}\rangle, ...\}$

   b. $\{..., Kermit, \ Mr.Toad, ...\}$

This claim – that *frog* expresses a function that has (2b) as its extension – says nothing about how the extension is determined. It says nothing about what makes an entity count or fail to count as a frog. And in principle, there might be various different ways of stating the necessary and sufficient conditions.

This point is perhaps easier to see with a mathematical example. Consider the functions in (3a) and (3b). For any value of x, both functions will yield the same output, the set of ordered pairs in (3c).

---

[1]Gritty is an orange creature (see tylerknowlton.com/images/gritty.jpg)

(3)   a. $F(x) = sin(x)$

      b. $F(x) = \frac{1}{2}ie^{(-ix)} - \frac{1}{2}ie^{(ix)}$

      c. $\{..., \langle \pi, 0 \rangle, \langle \frac{\pi}{2}, 1 \rangle, \langle 0, 0 \rangle, ...\}$

As Church (1941) put it, (3a) and (3b) name identical functions in extension, but they name different functions in intension. That is, while the outputs are identical, the procedure for getting from input to output differs.

Chomsky (1986) discussed this same distinction in the context of shifting the focus away from studying E(xternal)-language – the target of inquiry in structuralist linguistics and behaviorist psychology – and toward studying I(interal)-language (Pietroski, 2017). Whereas E-language refers to the set of sentences produced by a grammar, I-language refers to the system for producing them. As Chomsky and others stressed, the child's task in acquiring a grammar is not one of memorizing an infinite list of sentences or a set of speech behaviors, but one of internalizing a system for producing all and only the possible sentences of that language (see Lidz and Perkins (2018) and Lidz (2020) for helpful primers on language acquisition research explicitly situated in this framework).

In this context, imagine a child acquiring the meaning of *and*. Suppose they have a mind with two primitive concepts: conjunction (&) and "norjunction" ($\downarrow$). These operations give rise to the truth table in Table 1.1 (P & Q can be glossed P *and* Q; $P \downarrow Q$ can be glossed *neither P nor Q*). As the table makes clear, P & Q and $(P \downarrow P) \downarrow (Q \downarrow Q)$ are extensionally equivalent.

Given that our hypothetical learner has a primitive concept of conjunction and

P	Q	P&Q	P↓Q	P↓P	Q↓Q	(P↓P)↓(Q↓Q)
1	1	1	0	0	0	1
1	0	0	0	0	1	0
0	1	0	0	1	0	0
0	0	0	1	1	1	0

Table 1.1: Truth table for conjunction and norjunction

a primitive concept of norjunction, we could ask which concept they pair with the pronunciation of *and*. Do they learn that *P and Q* is an instruction to conjoin *P* and *Q* or that it is an instruction to norjoin *P* with itself, norjoin *Q* with itself, and norjoin the result? On its face, the answer seems obvious: Surely (4a) is a better description of the meaning of *and* than (4b).

(4)   a. *P & Q*

     b. $(P \downarrow P) \downarrow (Q \downarrow Q)$

But these only correspond to distinct hypotheses if we focus on the function in intension. Otherwise, (4a) and (4b) are just two different names for the same function and the choice between them is merely a notational one.

This state of affairs represents one version of a problem raised by Foster (1976): There are many distinct but equally true theorems that capture a given truth-condition, not all of which are equally plausible meaning specifications. If the expression *P and Q* is TRUE if and only if (4a), then it's TRUE iff (4b), and it's likewise TRUE iff $P \ \& \ Q \ \& \ 2 + 2 = 4$. But, as Larson and Segal (1995) put it, "is TRUE iff" and "means that" are two different relations. It certainly seems wrong to say that the expression *P and Q* means that $P \ \& \ Q \ \& \ 2 + 2 = 4$.

To take another example, consider two ways of specifying the meaning of the sentence in (5a):

(5)   a. Four frogs are green

      b. $|\{x : Green(x) \mathbin{\&} Frog(x)\}| = 4$

      c. $\exists w \exists x \exists y \exists z : Frog(w) \mathbin{\&} Frog(x) \mathbin{\&} Frog(y) \mathbin{\&} Frog(z) \mathbin{\&}$

          $Green(w) \mathbin{\&} Green(x) \mathbin{\&} Green(y) \mathbin{\&} Green(z) \mathbin{\&}$

          $(w \neq x) \mathbin{\&} (w \neq y) \mathbin{\&} (w \neq z) \mathbin{\&} (x \neq y) \mathbin{\&} (x \neq z) \mathbin{\&} (y \neq z) \mathbin{\&}$

          $\forall v : (Frog(v) \mathbin{\&} Green(v))[(w = v) \vee (x = v) \vee (y = v) \vee (z = v)]$

The more familiar (5b) reports the cardinality of the set of green frogs. The more cumbersome (5c) encodes the same content in a completely first-order way, by saying that there are four distinct individuals – w, x, y, and z – each of which is a frog, is green, and is not identical to any of the others; and that those four are the only individuals that meet these criteria (any v that's a green frog is either w, x, y, or z).

Abstracting away from the contributions of *frog* and *green*, we might say that *four* expresses a function that, given an internal predicate **I** (the first one it combines with, syntactically) and an external predicate **E** (the second one it combines with, syntactically), outputs TRUE just in case there are four things in the relevant domain that satisfy both predicates. We could then choose to specify that function as (6a) or (6b).

(6) $[\![four]\!] =$

      a. $\lambda\mathbf{I}.\lambda\mathbf{E}.|\{x : \mathbf{I}(x) \mathbin{\&} \mathbf{E}(x)\}| = 4$

b. $\lambda\mathbf{I}.\lambda\mathbf{E}.\exists w\exists x\exists y\exists z : \mathbf{I}(w) \;\&\; \mathbf{I}(x) \;\&\; \mathbf{I}(y) \;\&\; \mathbf{I}(z) \;\&$

   $\mathbf{E}(w) \;\&\; \mathbf{E}(x) \;\&\; \mathbf{E}(y) \;\&\; \mathbf{E}(z) \;\&$

   $(w \neq x) \;\&\; (w \neq y) \;\&\; (w \neq z) \;\&\; (x \neq y) \;\&\; (x \neq z) \;\&\; (y \neq z) \;\&$

   $\forall v : (\mathbf{I}(v) \;\&\; \mathbf{E}(v))[(w = v) \vee (x = v) \vee (y = v) \vee (z = v)]$

Both are extensionally equivalent: Given some domain, some **I** predicate, and some **E** predicate, both (6a) and (6b) will yield the same output.

But what does knowing the meaning of the lexical item *four* consist in? We could imagine many possible answers. It might be that what a speaker knows when they know the meaning of *four* is just the truth-conditional contribution it makes to a sentence. On this view, whether a particular speaker mentally represents (6a) or (6b) might be subject to random variation or to variation depending on the context. Maybe when a speaker learns the meaning of *four*, they learn to associate the pronunciation with an equivalence class of functions, all of which have identical extensions. This would make either choice – (6a) or (6b) – an equally good theoretical description of how speakers understand *four*.

Another possibility is that speakers' representations of *four* share more in common than expressing the same truth-conditional content. It's not implausible to think that (6b) is far too cumbersome for humans to lexicalize. Maybe then, every competent English speaker has a concept like (6a) connected with the pronunciation of *four* and it is only after taking a logic class that we come to recognize something like (6b) as extensionally equivalent. At the same time, we could imagine a mind without any concept of cardinality. Such a mind would be unable to represent (6a),

but could still represent (6b). Speakers with both minds could learn a word that we might translate as *four* and they would never disagree in conversation about, say, whether there were four green frogs. If either one of these possibilities turned out to be true, (6b) and (6a) would not be equally good theoretical descriptions of how a certain mind understands *four*.

The above example illustrates the relevant sense in which formal semantic claims about lexical meanings can be treated as psychological hypotheses. Namely, instead of treating a lambda expression like (6a) as a theorist's choice about how to write down a specification of some function in extension, we can treat it as a representation of a particular function in intension, which might share an extension with any number of other functions. Then we can ask whether one of these "psycho-logical forms" is a better description of the expression as represented in the minds of speakers. And we can say that a formula is a better description than alternatives to the extent that it explicitly encodes the operations that are necessarily mentally represented by speakers in the course of mentally representing the expression.

This leaves open the possibility that the right "psycho-logical form" of some grammatically atomic expression might itself be atomic (Fodor, 1998). It might be, for example, that there is a primitive concept FOUR – which has no internal structure – that serves as the meaning of the expression *four*. But, in the more interesting cases, proposals about representational format will involve some amount of semantic decomposition. That is, they will involve claims of a grammatically atomic expression being represented in terms of multiple (inaudible) parts (e.g., (6a) involves notions of cardinality and equality; (6b) involves existential quantification,

conjunction, non-identity, universal quantification, and identity).

Williams (2015) helpfully distinguishes three flavors of semantic decomposition. The weakest decompositional claim is metasemantic: The formula is merely a way for a theorist to regiment a meaning. As discussed above, this is how the differences between extensionally equivalent specifications, like (6b) and (6a), are usually understood. Choosing one over the other may make it easier for the theorist to state a generalization, but the choice is a notational one and it does not correspond to any sort of psychological hypothesis.

A stronger decompositional claim is a representational one: The meaning of the expression itself does not decompose, but representing it does routinely trigger representation of other concepts. For example, imagine that the meaning of *frog* is just (an instruction to access) the atomic concept FROG but that every time you token the concept FROG, you also token the concept ANIMAL. Then we might claim that while the meaning of *frog* itself does not have any internal structure, there is still some sense in which the lexical item *frog* has ANIMAL as a constituent part.

A stronger claim still is one of strict semantic decomposition: The meaning is made up of smaller parts that are not themselves expressed. For example, imagine that not only does *frog* routinely cause ANIMAL to be tokened, but that the meaning of the expression genuinely is built, in part, from the ANIMAL concept. The claim here is that representing the expression *frog* necessarily requires, among other things, tokening the concept ANIMAL. It would not be possible, given strict semantic decomposition, to understand the meaning of *frog* without having the AN-IMAL concept. This last flavor, strict decomposition, is the relevant one for thinking

11

about an expression's representational format. But it is also the hardest to support empirically (by a large margin), as discussed below.

Before getting to possible sources of evidence though, it is important to note that questions about representational format along these lines can be posed independent of questions about how meanings compose. The compositional semantics is unchanged between, for example, (6a) and (6b). So we can still say, in keeping with standard semantic theory, that *four* has a meaning that combines with the meanings of two predicates (**I** and **E**), both of which are supplied by function application. What changes between (6a) and (6b) is what comes after the lambda expressions. As Wellwood (2020) puts it, we can think of these representations as tracking "morphosyntactic structure on the left" of the lambdas and tracking "conceptual structure on the right" (p.8).

Since this dissertation will have nothing to say about the compositional semantics of universal quantifiers, we can remain agnostic about whether having lambdas "on the left" is the right way to track morphosyntactic structure. It is certainly not the only possibility.[2] Instead, the focus of this dissertation will only be what happens "on the right side" of lambda expressions. So when lambdas appear in proposed representations, they are not intended to be read as part of the psychological claim (though what comes after "$\lambda\mathbf{I}.\lambda\mathbf{E}$." is intended to be read as a claim about the meaning's representational format).

_____

[2] Alternatively, we might adopt a view that dispenses with truth-values and avoids treating predicates as relations; see Chapter 7 of Pietroski (2018). And instead of a categorematic specification for *each* like $\lambda\mathbf{I}.\lambda\mathbf{E}.\forall x(\mathbf{I}x)[\mathbf{E}x]$, we might propose a syncategorematic rule for how truth-conditions of sentence frames like $[_S [_D \ each \ [_N ...]]_i \ [_S ...t_i..]]$ are specified; see section 2.1 of Chapter 2. The same questions about representational format arise regardless.

1.2 Probing representational format linguistically

Claims that expressions have parts that are silent but meaningful are not uncommon in semantics (see Engelberg (2011) for a helpful review). They have been popular in work on causative verbs (e.g., Lakoff 1966; Fillmore 1968; Rappaport Hovav and Levin 1998; Pietroski 2003b; Levin and Rappaport Hovav 2005; though see Fodor 1970), event semantics (e.g., Davidson 1967a; Parsons 1989; Kratzer 1996; Pietroski 2005), and quantification (e.g., Geurts and Nouwen 2007; Hackl 2009; Pietroski et al. 2009; Lidz et al. 2011), to name a few places.

Some proposals resolve semantic decomposition syntactically. That is, they treat what appears on the right side of lambda expressions as atomic, but decompose the "word" in question into smaller units of syntactic composition. For example, Zeijlstra (2011) argues that the German negative quantifier *keine* decomposes into a negative operator and an indefinite, which are syntactically separable. In particular, a modal can scope between the negative operator and the indefinite, suggesting that *keine* is semantically and syntactically complex.[3]

But syntactic arguments of this sort cannot be made to support strict semantic decomposition of some (syntactically atomic) lexical item. Providing evidence for this sort of decomposition – evidence that an expression with no syntactic parts nonetheless has a meaning with structure – has proven notoriously difficult. The main reason is that most of the non-syntactic generalizations can be captured by

---

[3] The relevant examples are sentences like *du must keine Krawatte anziehen* ≈ "you must wear no tie", which can give rise to the "split-scope" reading: *It is not required that you wear a tie* ($\neg > \Box > \exists$).

appealing to the weaker notion of representational decomposition. To borrow an example from Williams (2015), since ewes are female sheep, anything that falls under the concept EWE also falls under the concept FEMALE. Accordingly, the inference from (7a) to (7b) is compelling.

(7) a. Eunice is a ewe

 b. Eunice is female

We might describe this by saying the truth of (7b) is entailed by the truth of (7a). This observation makes it tempting to suppose that *ewe* has as its meaning a structured representation made up of the concepts FEMALE and SHEEP (perhaps among other things). This would put the inference from (7a) to (7b) on par with the inference from (8a) to (8b), which holds in virtue of the logical structure alone.

(8) a. Eugene is a white sheep

 b. Eugene is a sheep

But the claim that the lexical item *ewe* is in part composed of the concept FEMALE would not be justified, since the inference from (7a) to (7b) might reflect what we know about ewes without reflecting the meaning of *ewe*. As Fodor and Lepore (1998) put it, echoing Quine, "What distinguishes linguistic knowledge from world knowledge? What distinguishes lexical entailment from mere necessity?" (p.274).

These same questions can be asked about any case of proposed lexical decomposition, at least for common nouns. Bachelors are necessarily unmarried, but does the lexical item *bachelor* have the concept UNMARRIED as a constituent? Not

necessarily. Especially if the null hypothesis is that meanings are atomic (i.e., un-structured), what seem like analytic inferences (*Kermit is a bachelor* → *Kermit is unmarried*) might be accommodated by appealing to strongly associated world knowledge instead of lexical decomposition (Fodor et al., 1975). It might be that people know that bachelors are unmarried, independently of being able to under-stand the lexical item.

This argument – call it "Fodor's challenge" – urges caution when probing representational format linguistically. For any proposed semantic decomposition (of a syntactically atomic lexical item), retreating to world knowledge is always a logical possibility. But as the examples in sections 1.2.1 and 1.2.2 will illustrate, there may be some inferential patterns that are more compelling than *ewe* → *female* and *bachelor* → *unmarried*.

1.2.1 Action sentences and compelling inferences

Action sentences potentially provide one such example of compelling infer-ences. Davidson (1967a) was concerned with the patterns of inferences that can be made between related sentences, like those in (9).

(9)   a. Kermit watched baseball with Gritty in Philly

   b. Kermit watched baseball with Gritty

   c. Kermit watched baseball in Philly

   d. Kermit watched baseball

In particular, these sentences give rise to what later came to be known as the diamond pattern of inferences in (10).

(10) (9a) $\longrightarrow$ (9b) & (9c)

$$(9b) \qquad (9c)$$

$$(9d)$$
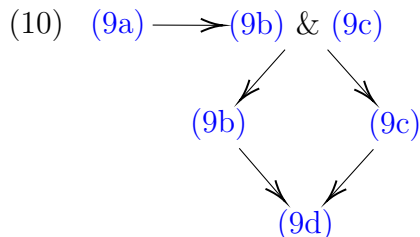
From (9c), for example, it follows that (9d): If Kermit watched baseball in Philly, then Kermit watched baseball.

Davidson pointed out that treating sentence like those in (9) as distinct two-place relations, like those in (11), fails to explain the generalization about inferences stated in (10).

(11)   a. *Watched-with-Gritty-in-Philly*($Kermit$, $baseball$)

 b. *Watched-with-Gritty*($Kermit$, $baseball$)

 c. *Watched-in-Philly*($Kermit$, $baseball$)

 d. *Watched*($Kermit$, $baseball$)

Why should the relation in (11a), for example, bear any particular relationship to, say, the relation in (11d)? And do we really understand the sentences in (9) as distinct relations?

Instead, he proposed treating (9a) more like (12), where *watched* expresses a three-place relation between an event e, the watcher (*Kermit*), and the thing watched (*baseball*). The adverbial modifiers – *with Gritty* and *in Philly* – are treated as separate conjuncts.

16

(12) $\exists e[Watched(e,\ Kermit,\ baseball)\ \&\ withGritty(e)\ \&\ inPhilly(e)]$

$\approx$ *There was a watching of baseball by Kermit and it was with Gritty and it was in Philly*

Given this sort of three-place relation, getting from (9a) to the other sentences in (9) is just a matter of eliminating conjuncts. This explains the generalization in (10) as a matter of logical syntax, much like *white sheep* $\rightarrow$ *sheep*.

Fodor's challenge looms large. As it was with common nouns, it might be that *watch* has as its meaning an unstructured representation WATCH and that we know, independent of language, that watchings with someone are watchings. But unlike the situation with common nouns, the network of inferences in (10) do not appear to be specific to world knowledge (compare them to *Kermit watched baseball* $\rightarrow$ *A frog watched baseball*, which relies on knowing that Kermit is a frog). As Pietroski (2005) puts it, conditionals like *Kermit sang if Kermit sang loudly* "seem to be risk-free in a way that other conditionals are not" (p.15). These inferences remain compelling even if the particular lexical items are changed. And it is hard to imagine a speaker knowing the meaning of *watch* without knowing that the pattern in (10) holds.[4] That makes the inference from (9a) to (9b), (9c), or (9d) feel more like the inference from *white sheep* to *sheep*, which even Fodor agrees is compelling in virtue of logical structure alone.

Since Davidson (1967a), there have been many advances in event semantics

---

[4] In fact, the inferences remain compelling even if one does not know the meaning of *watch*. For example, if *watch* were replaced with a nonce word, the inference from *Kermit gleebed baseball in Philly* to *Kermit gleebed baseball* seems on a par with the inference from *Kermit is a green florp* to *Kermit is a florp*.

(e.g., Parsons 1989; Kratzer 1996; Landman 2000; Pietroski 2005; Schein 2012; Williams 2021). For example, instead of treating *watch* as a three-place predicate, as in (12), at least AGENT, and sometimes also PATIENT, are severed from the event predicate, yielding representations more like (13) (see section 1.4.2).

(13) Kermit watched baseball

$\exists e[Watched(e) \ \& \ \text{AGENT}(e, Kermit) \ \& \ \text{PATIENT}(e, baseball)]$

And, to be sure, this structure might not reflect decomposition of the verb itself. It might not be that *watch* decomposes into something with AGENT as a constituent; AGENT might instead reflect the contribution of a covert syntactic operator (e.g., Kratzer 1996). Davidson's own position was that the structure in (12) is syntactically derived (because each conjunct is syntactically separate from the rest of the clause).

Regardless of the specifics, the point here is just that a retreat to strongly associated world knowledge, though logically possible, threatens to miss a generalization about inferences that speakers find compelling. So these sorts of generalizations can offer initial – though far from dispositive – evidence for a particular representational format.

It is worth noting that Davidson himself did not go so far, as he was not concerned with the psychology of speakers. His aim was to show that, as a theorist, truth-conditions of action sentences can be stated in a way that makes entailments follow for principled reasons. The proposed decomposition was therefore intended to be metasemantic: It was meant as a way for the theorist to state generalizations,

not as a claim about speakers' mental representations. Whether this was the right approach is related to whether the facts in question are better thought of as generalizations about what sorts of sentences preserve truth in a model or as generalizations about what sorts of inferences speakers naturally find compelling (see e.g., Chapter 5 of Pietroski (2018) for discussion).

1.2.2 Complex superlative quantifiers and inference interference

Geurts and Nouwen (2007) provide another example of using linguistic evidence to argue for one particular representational format over another.[5] And unlike the case of transitive verbs considered above, the proposed decomposition here is not syntactically resolved. Geurts and Nouwen's main concern is the meanings of composite superlative quantifiers like *at least n* or *at most n* and composite comparative quantifiers like *more than n* or *fewer than n*. We might be tempted to specify these complex quantifiers in terms of notions like GREATER-THAN and LESS-THAN (as Keenan and Stavi (1986) and others since them have). For example, consider the representations in (14),[6] which only differ in whether they use the symbol "$\geq$", "$>$", "$\leq$", or "$<$".

---

[5] They never explicitly state that the target of inquiry is representational format, but Geurts et al. (2010) summarize their view by saying: "Contrary to the standard model, G&N's analysis predicts that the logic of quantifiers like *at least/most 3* is different from that of quantifiers like *more/fewer than 3*" (p.131). We might understand "logic" here to mean natural logic or "psychologic". They later go on to say that "The principal tenet of G&N's theory is that the meanings of quantifiers are not homogeneous: even if they appear to be closely related, the meanings of superlative quantifiers are essentially richer than those of their comparative counterparts, and not expressible in basic set theory" [i.e., as instances of generalized quantifiers] (p.147).

[6] In (14), the second-order variable X can either be thought of as ranging over plural entities, like sets, or as ranging over individuals and taking on more than one value at a time; see Chapter 3 for discussion of the distinction. In any case, the second-order X is meant to be contrasted with the first-order x, which can be thought of as ranging over individuals and as taking on only a single value at a time.

(14) a. At least n frogs are green

$\exists X[Frog(X)\ \&\ Green(X)\ \&\ |X| \geq n]$

$\approx$ *There exist some things$_X$ s.t. they$_X$ are frogs and they$_X$ are green*

and they$_X$ number n or more

b. More than n frogs are green

$\exists X[Frog(X)\ \&\ Green(X)\ \&\ |X| > n]$

$\approx$ *...and they$_X$ number more than n*

c. At most n frogs are green

$\exists X[Frog(X)\ \&\ Green(X)\ \&\ |X| \leq n]$

$\approx$ *...and they$_X$ number n or fewer*

d. Fewer than n frogs are green

$\exists X[Frog(X)\ \&\ Green(X)\ \&\ |X| < n]$

$\approx$ *...and they$_X$ number fewer than n*

Psychologically speaking, we might suspect that these symbols correspond to primitive concepts like GREATER-THAN and LESS-THAN.

In any case, given the representations in (14), we can see why relationships hold between the two sentences in (15a) and the two sentences in (15b).

(15) a. At least 3 frogs are green $\leftrightarrow$ More than 2 frogs are green

$|X| \geq 3 \leftrightarrow |X| > 2$

b. At most 2 frogs are green $\leftrightarrow$ Fewer than 3 frogs are green

$|X| \leq 2 \leftrightarrow |X| < 3$

But Geurts and Nouwen point out that superlative and comparative quantifiers do not pattern together with respect to all patterns of inference.

For example, from (16) it follows that (17a), but it seems less obvious that (17b) follows.

(16) Kermit ate exactly three hot dogs

(17)   a. Kermit ate more than two hot dogs

       b. Kermit ate at least three hot dogs

Judgments here are subtle, but Geurts et al. (2010) support them with experimental semantics evidence. Their data suggest that, at least, the inference (16) → (17a) and (16) → (17b) are not on a par. What seems to be at issue in the latter case is that (17b) more strongly suggests the possibility that Kermit had more than three hot dogs, which we know to be false from (16). This possibility is consistent with (17a) as well – both (17b) and (17a) entail that it's possible that Kermit ate four hot dogs – but *more than two* doesn't seem highlight this possibility in the same way. As a consequence, the inference (16) → (17b) is not as compelling as we might expect it to be.

The issue cannot just be that *three* is mentioned in (17b) but not (17a), because the same effect arises in (18a) and (18b), neither of which mention the actual number of hog dogs Kermit ate.

(18)   a. Kermit ate fewer than five hot dogs

       b. Kermit ate at most four hot dogs

In particular, from (16) it clearly follows that (18a), but it less clearly follows that (18b). Again, both (18a) and (18b) entail the possibility that Kermit ate four hot dogs, but somehow, (18b) foregrounds this possibility more strongly. And by foregrounding this possibility, the inference from (16) to (18b) is impeded.

To account for this pattern of inferences (along with some additional data not yet discussed), Geurts and Nouwen propose doing away with "$\leq$" and "$<$" altogether and, instead of treating "$\geq$" as an atomic symbol, separating it into "$>$" and "$=$". They then propose introducing "$= n$" and "$> n$" in two separate conjuncts. Importantly, each conjunct is said to have a different epistemic modal status. For *at least*, the "$= n$" conjunct must necessarily hold, whereas the "$> n$" conjunct is possible, but not necessary. For *at most*, the "$= n$" is possible, but not necessary, whereas the "$> n$" conjunct is not possible.

So instead of the specifications in (14), this yields specifications more like (19), where the complex superlative quantifiers *at least n* and *at most n* are treated as modal expressions with two separate conjuncts and (less importantly) neither *at most n* nor *fewer than n* are specified in terms of a LESS-THAN concept, like "$<$".

(19)   a. At least n frogs are green

$\Box \exists X [Frog(X) \& Green(X) \& |X| = n] \&$

$\Diamond \exists Y [Frog(Y) \& Green(Y) \& |Y| > n]$

$\approx$ *It's necessary that there exist some things$_X$ s.t. they$_X$ are frogs and they$_X$ are green and they$_X$ number n, and it's possible that there exist some things$_Y$ s.t. ...they$_Y$ number more than n*

b. More than n frogs are green

$\exists X[Frog(X)~\&~Green(X)~\&~|X| > n]$ (same as (14b))

$\approx$ ...and they$_X$ number more than n

c. At most n frogs are green

$\Diamond \exists X[Frog(X)~\&~Green(X)~\&~|X| = n]~\&$

$\neg \Diamond \exists X[Frog(X)~\&~Green(X)~\&~|X| > n]$

$\approx$ *It's possible that there exist some things$_X$ s.t. they$_X$ are frogs and they$_X$ are green and they$_X$ number n, and it's not possible that there exist some things$_Y$ s.t. ...they$_Y$ number more than n*

d. Fewer than n frogs are green

$\neg \exists X[Frog(X)~\&~Green(X)~\&~|X| = n]$

$\approx$ *There do not exist some things$_X$ s.t. they$_X$ are frogs and they$_X$ are green and they$_X$ number n*

Given these representations, the superlative *at least n* and *at most n* both make explicit the possibilities that seemed to lead to the problematic inferences above. Namely, on this view, *Kermit ate at least three hot dogs* is represented in a way that explicitly encodes the possibility that Kermit ate more than three, whereas *Kermit ate more than two hot dogs* is not. And likewise, *Kermit ate at most four hot dogs* is represented in a way that explicitly encodes the possibility that Kermit ate four, whereas *Kermit ate fewer than five hot dogs* does not. So while all representations entail the problematic possibility, only the complex superlative quantifiers make it explicit in the representation.

The idea that *at least* and *at most* have modal meanings may also help explain their distributional restrictions. In particular, complex comparative quantifiers, but not complex superlative quantifiers, can appear under the scope of negation, as seen in (20).

(20)   a. Kermit didn't have {?at least three / more than two} hot dogs

         b. Kermit didn't have {?at most three / fewer than four} hot dogs

This makes superlatives similar to some epistemic modals, like *might* and *perhaps*, which also resist scoping under negation, as in (21).

(21)   a. Kermit might not have had three hot dogs $\diamondsuit > \neg;\ ?\neg > \diamondsuit$

         b. ?Kermit didn't have perhaps three hot dogs

Geurts and Nouwen (2007) and Geurts et al. (2010) provide other sources of data supporting the proposed specifications in (19). Though, to be sure, it remains a possibility that the standard specifications in (14) are correct and data pointing to the contrary can be accommodated pragmatically (see Coppock and Brochhagen (2013a,b) for one such proposal). If a pragmatic approach turns out to be right, then the data reviewed here will ultimately not provide evidence for complex superlative quantifiers having a meaning with one particular representational format over another.

Of course, the aim of this section has not been to take a hard line in favor of Geurts and Nouwen's account or any other, but to offer another example of how linguistic evidence (e.g., inferential and distributional) might be leveraged to probe

a meaning's representational format. Having seen these linguistic examples, we now turn to potential psycholinguistic sources of evidence.

1.3 Probing representational format psycholinguistically

As with traditional linguistic evidence, psycholinguistic evidence for strict semantic decomposition has been hard to come by. For one thing, Fodor's challenge needs to be met by any experiment purporting to provide evidence for decomposition. Imagine an experiment – perhaps using a futuristic brain imaging technique – showing that every time a participant processes the expression *Kermit is a bachelor*, they token the concept UNMARRIED. This might reflect the lexical item *bachelor* decomposing into ...UNMARRIED MAN..., but it also might reflect a strong association between thoughts about bachelors and the knowledge that they are all unmarried.

Put another way, it wouldn't be all that surprising to find out that people think UNMARRIED when they process an expression that contains *bachelor*. If you grew up in America watching football, you probably can't avoid thinking about Sunday when someone mentions the Super Bowl (and if you're only a causal observer of American football, maybe you can't avoid thinking about advertisements when someone mentions the Super Bowl). But this alone hardly suggests that the expression *superbowl* has the concept SUNDAY (or ADVERTISEMENTS) as part of its meaning. It might just be that *superbowl* has SUPERBOWL as its meaning and individual speakers strongly associate that concept with some others, perhaps including SNACKS, ADVERTISEMENTS, FEBRUARY, and, in my case, PHILADELPHIA-EAGLES.

To make matters worse, experiments in this area have been largely unsuccessful. As Fodor (1998) puts it: "It's an iron law of cognitive science that, in experimental environments, definitions always behave exactly as though they weren't there" (p.46). By "definitions," Fodor has in mind strict semantic decomposition of the sort at issue here. And the experimental failures motivating this iron law do not involve futuristic brain scans, but participants being shown, via behavioral measures, to process terms like *bachelor* with no more difficulty than terms like *not married*. In one such case, Fodor et al. (1975) asked participants to evaluate arguments like (22).

(22) If practically all of the men in the room are {not married / bachelors}, then few of the men in the room have wives

Participants were faster to judge (22) (and sentences like it) as a valid or invalid argument given *bachelor* than given *not married*. This would be surprising, the reasoning goes, if *bachelor* decomposed into the concepts NOT MARRIED and MAN.[7] Fodor et al. (1980) report similar failed experiments that do not use reaction time as the dependent measure.

In light of these concerns, strong psycholinguistic evidence for a particular representational format must satisfy at least two criteria. First, it must try to meet Fodor's challenge. Given that a retreat to associated world knowledge is always a logical possibility, meeting this challenge is difficult. Clearly, evidence that rep-

---

[7] Fodor (1998) notes that it could still be true that the decomposition gets "compiled" in a way that would preclude reaction time experiments from detecting it. He suggests this possibility is too post-hoc to be taken seriously. But, more charitably, if increased processing difficulty was the wrong linking hypothesis in the first place, then these experiments had no hope of revealing underlying semantic structure.

resenting a given expression leads to representing a prominent entailment of that expression (e.g., that representing the expression *bachelor* leads to tokening the concept UNMARRIED) could straightforwardly be predicted by association alone. And to the extent that association alone predicts a result, it is unable to meet Fodor's challenge. This makes open-class lexical items like common nouns (and, to a lesser extent, causative verbs) more difficult cases to consider. As we will see in section 1.3.2, experiments probing the logical vocabulary (e.g., determiners) provide results that do a better job of resisting explanation in terms of association.

Second, given past experimental failures leading to Fodor's iron law, the linking hypothesis from a decompositional claim to the supporting data must be more sophisticated than the idea that greater complexity leads to greater processing difficulty. Section 1.3.1 presents an alternative that will be used to motivate the experiments on *most* reported in section 1.3.2 and the experiments on the universal quantifiers reported in Chapters 2 and 3.

1.3.1 Interface transparency as a linking hypothesis

In terms of a more sophisticated linking hypothesis about how semantic representations are related to observable behavior, Lidz et al. (2011) propose (23).

(23) Interface Transparency Thesis (ITT): 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, all else equal, people will evaluate a sentence with a procedure that transparently reflects that sentence's meaning. If meanings are thought of as providing instructions to conceptual systems, then the thought behind the ITT is that the "path of least resistance" is to follow that instruction as it's written.

This is not to say people will always use a certain strategy to evaluate a given sentence. In ordinary (i.e., uncontrolled) situations, there are many factors that can be relevant for choosing a verification strategy. The claim of the ITT is just that the details of the semantic representation carry some weight in determining the strategy used and that its weight can be detected in carefully controlled contexts (see Pietroski et al. (2011) and Hunter et al. (2017) for more discussion on this point). Only when other considerations are held equal can inferences about the expression's semantic representation be made from evidence about verification procedures naturally deployed to evaluate it.

This idea is in the spirit of Marr (1982), who discussed how the representational format of a given thought can highlight the applicability of certain computational procedures and background others. To borrow Marr's example, THIRTY SEVEN can be represented by some computing device using the Arabic (24a) or the binary (24b).

(24)   a. 37

        b. 100101

As with the examples discussed in section 1.1, both of these alternatives have identical content, but that content is represented in a different format. While (24a) makes decomposition into powers of 10 explicit, (24b) makes decomposition into powers of

2 explicit. As a consequence, the Arabic system makes it easy to know whether a number is a power of 10: Just check to see if there is a 0 in the 10^0 position. At the same time, whether the number in question is a power of 2 is not directly encoded in the representation, so further computations are required. The situation is exactly reversed for the binary system: Figuring out whether the number is a power of 2 is easy – just check to see whether there is a 0 in the 2^0 position – figuring out whether it is a power of 10 takes some work.

In the same way, the idea behind the ITT in (23) is that certain ways of evaluating the thought built by executing an expression's meaning will be more natural than others (Pietroski et al., 2011). So if evidence suggests that participants routinely use strategy$_1$ to evaluate some expression even when strategy$_2$ is cognitively available, one explanation is that strategy$_1$ is a more transparent reflection of the expression's meaning. Better still, if participants use strategy$_1$ to verify expression$_1$ and use strategy$_2$ to evaluate the minimally-different expression$_2$, this variation in verification strategy can be reasonably attributed to a difference in meaning between expression$_1$ and expression$_2$. And if these expressions make identical truth-conditional contributions, then the result can be reasonably attributed to a difference in representational format. This is exactly the situation in the example discussed in section 1.3.2 below.

Before getting to that example, a caveat is in order. Section 1.3 questioned the linking hypothesis underlying the failed experiments Fodor (1998) used for the basis of his iron law, and the ITT in (23) can be questioned just the same. Given any result, we find ourselves in the unfortunate position of having to use that result

29

to defend the linking hypothesis (that meaning and verification are related in the way described above) and the lexical semantic hypothesis (that an expression is represented in one particular format as opposed to another).

But being in this position is not unique. As Fodor (1998) reminded readers, in the context of conceding that no one has proven that there aren't any definitions: "Cognitive science doesn't do proofs; it does empirical, non-demonstrative inferences" (p.46). And any inference from empirical findings requires a linking hypothesis. Some linking hypotheses have been around for a while, and consequently have received a good deal of empirical support (e.g., duration of eye gaze as an index of processing difficulty; acceptability as an index of grammaticality). Others, like the ITT, are more nascent. But like any linking hypothesis, (23) will ultimately be vindicated to the extent that it leads to the discovery of various phenomena as opposed to a single cul-de-sac of results.

The case study on *most* described in section 1.3.2 provides an example of one such phenomenon, and the remainder of this dissertation, focusing on the universal quantifiers, presents another. For others in the same vein, see Odic et al. (2018) on the count/mass distinction and Wellwood (2020) on degree semantics.

1.3.2   The case study of *most*

As with the examples from section 1.1, the meaning of an expression like *most frogs are green* can be specified in many ways, including those in (25).

(25)   a. $|\{x : Frog(x) \ \& \ Green(x)\}| > \frac{1}{2}|\{x : Frog(x)\}|$

b. $|\{x : Frog(x) \ \& \ Green(x)\}| > |\{x : Frog(x) \ \& \ \neg Green(x)\}|$

c. $|\{x : Frog(x) \ \& \ Green(x)\}| > |\{x : Frog(x)\}| - |\{x : Frog(x) \ \& \ Green(x)\}|$

d. $OneToOnePlus(\{x : Frog(x) \ \& \ Green(x)\}, \{x : Frog(x) \ \& \ \neg Green(x)\})$

The first three representations, (25a) - (25c), specify *most* in terms of cardinality.[8]
The "more than half" specification in (25a) explicitly encodes the proportion of frogs
that must be green. In contrast, the "negation" version in (25b) relies on predicate
negation and the "subtraction" version in (25c) instead relies on representing all
frogs (in the relevant domain) and subtracting the green ones. Lastly, the "one-to-
one-plus" specification in (25d) says that the green frogs and the non-green frogs
correspond one-to-one with at least one leftover green frog. So *most* can be specified
with or without a proportion like $\frac{1}{2}$, with or without predicate negation, and with
or without appeal to cardinality. The question is whether one of these specifications
is a better description of how speakers mentally represent *most*.

Note that the question of how *most* is specified, as framed above, is a question
of semantic decomposition. This does not preclude the possibility that *most* is also
subject to syntactic decomposition. For example, Bobaljik (2012) provides cross-
linguistic evidence for the idea that superlatives decompose along the following lines:
[[[ADJ] -ER] -EST]]]. On this view, *most* would not be a syntactic atom but might
syntactically decompose into something like (26).

_____

[8] For ease of exposition, the specifications in (25) assume combination with a count noun, like
frog or *dot*. See Odic et al. (2018) for details concerning mass nouns. In general, the bars || in
$|\{x : \mathbf{P}(x)\}|$ might represent a more general notation for measurement that serves as an instruction
to the Approximate Number System when the predicate $\mathbf{P}$ is supplied by a count noun and to the
Approximate Area System when the predicate $\mathbf{P}$ is supplied by a mass noun. So speaking of the
specifications implicating cardinality can be understood as shorthand for "implicating cardinality
or a measure of continuous extent, depending on the presence of a count/mass feature."

(26) [[[MANY] -ER] -EST]]]

One way to bring these two claims – syntactic and semantic decomposition – into alignment is to assume that not all syntactic atoms have an interpretation, just as not all syntactic atoms have a pronunciation (Preminger, 2020).[9] On this view, (26) may be a syntactically complex "LF formative" that has a meaning, even if its elements do not each, on their own, have a semantic interpretation. Another way would be to resolve some of the proposed semantic decomposition syntactically. If it turns out that one of the representations in (25a) - (25c) are the right specification of *most*, maybe the symbol ">" is supplied by the syntactic atom -ER. The point here is just that the two claims are not mutually exclusive.

As the rest of this section will only be concerned with semantic decomposition, we can remain agnostic about the possibility of syntactically decomposing *most*. The same is true for the rest of this dissertation when discussing the proposed semantic decompositions of *each*, *every*, and *all*. That said, for the universal quantifiers, there have been few – if any – claims of syntactic complexity. Therefore, if we find evidence of structured meaning representations, we likely have evidence for pure (i.e., not syntactically derived) strict semantic decomposition.

With this codicil out of the way, we can return to the possible specifications of *most* in (25). To get at the question of which of these possibilities is a better

---

[9] As evidence for this architectural claim, Preminger points to cases of partial overlap between pronunciation and interpretation like PAST GO OFF (i.e., *went off*). On the meaning side, GO and OFF have an idiomatic interpretation ("explode" or "be triggered") to the exclusion of PAST. But on the pronunciation side, PAST and GO get pronounced (as *went*) to the exclusion of OFF. So it need not be the case that each syntactic atom has an interpretation and a pronunciation. Instead, some combination of syntactic atoms serve as pronounceable "PF formatives," whereas potentially different combinations of the same syntactic atoms serve as meaningful "LF formatives."

description of how speakers understand *most*, Pietroski et al. (2009) asked participants to verify sentences like *most of the dots are blue* with respect to various types of blue and yellow dot-displays. On most display types, participants showed signs of using a cardinality-based strategy. That is, their ability to correctly answer was dependent on the ratio of blue dots to yellow dots (see section 2.2.2 in the following chapter for a review of the Approximate Number System).

Importantly, participants used cardinality even given displays where the blue and yellow dots were paired, making it easy to identify any leftover dots. To confirm the ease of identifying leftovers, participants were shown these same displays and asked to *find the loaner dot*. Sure enough, participants were better at finding leftover blue dots (when explicitly asked to do so) than they were at determining whether most of the dots were blue (when shown the same display and asked to evaluate the *most*-sentence). So a one-to-one correspondence strategy is psychologically available and would lead to better performance. Even so, participants opt for an inferior cardinality-based strategy.

These results point to *most* being specified in terms of cardinality (as in (25a) - (25c)), not correspondence (as in (25d)). If it were specified in terms of correspondence, the reasoning goes, more would need to be said to explain why participants avoid using a cognitively-available correspondence-based strategy to evaluate *most*-sentences in favor of using a sub-optimal cardinality-based strategy.

As noted in section 1.3.1, this conclusion is not challenged by showing that participants use other (non-cardinality-based) strategies to evaluate *most*-sentences in other contexts. Nothing about the hypothesis that *most* has a particular representa-

tional format precludes participants from using strategies that do not transparently reflect that representation. Sometimes, non-linguistic pressures urging a particular verification strategy are too great to ignore.

Pietroski et al. (2009) report and discuss one such case: When blue dots were lined up on one side of the screen and yellow dots were lined up on the other, participants used line length as a proxy for whether most of the dots were blue. If the blue line was longer, they answered TRUE; if the blue line was shorter, they answered FALSE. But because line length can be estimated far more accurately than cardinality, nothing can be concluded about the meaning of *most* from this result. On these trials, participants were wisely resorting to an easy and accurate strategy. If it had turned out that participants successfully used line length when asked to evaluate *blue is longer than yellow* but nonetheless used cardinality when asked to evaluate *most of the dots are blue*, this would have provided more evidence in favor of a cardinality-based specification. Conversely, if it had been the case that line length computations were inferior to cardinality computations and participants nonetheless used line length, then this result would tell against *most* being specified in terms of cardinality. But as it stands, this particular condition turned out not to be of any importance for claims about *most*'s representational format.[10]

To put the point more generally: The ITT contends that the representational

---

[10] The condition was worth running because we often don't know in advance which strategies will be superior but nonetheless dispreferred. Finding the "sweet spot" where the weight of the representational format can be detected is difficult. Moreover, the finding that participants can use line length as a proxy for numerosity is interesting in the context of questions about what sorts of verification procedures people can use and how flexibly they can shift between them. It might be that there are some superior verification procedures – e.g., one-to-one correspondence – that participants cannot be pressured to adopt by altering properties of the visual display.

format carries some weight in determining the verification procedure used, but it is obviously not the only consideration. If you don't speak English and someone asks you to verify whether *most of the dots are blue*, you might do well to adopt a verification strategy of asking your English-speaking friend to tell you the answer. If, in a crude experiment, the statement was TRUE whenever the blue dots appeared on the left and FALSE whenever the blue dots appeared on the right, you might do well to adopt a verification strategy that relied on the left/right distinction. From these cases we would conclude nothing about the representational format of *most*. The interesting cases are the ones in which participants avoid using a cognitively-available strategy that would lead to better responses in favor of a strategy that is sub-optimal but more closely aligns with a plausible representational hypothesis.

With that in mind, Pietroski et al. (2009) provide strong evidence for ruling out a correspondence specification like (25d). Taking this conclusion as a starting point, Lidz et al. (2011) extend the paradigm to argue against a negation specification like (25b). Participants in their task were asked to evaluate sentences like *most of the dots are blue*, but this time, displays contained up to five colors. Spelling out the predictions of the negation view is a bit more complex than it was for the correspondence view. Though the visual system can isolate a single color even in displays with many colored dots, it cannot direct attention to a heterogenous group, like the non-blue dots (see Wolfe (1998) for a helpful review of visual search). In a slogan: Negation of a visual feature is not itself a visual feature. So (25b) does not transparently map onto a possible strategy that treats the set defined using predicate negation as a visual ensemble defined by negation of a visual feature.

Instead, a participant wanting to enumerate the non-blue dots would have to do so by adding the estimated cardinality of each non-blue group (e.g., $\#(yellowDots)+\#(greenDots)+\#(redDots)+...$). Given just blue and yellow dots, this strategy is perfectly reasonable. In fact, as discussed below, it is the superior strategy. But even adding a third color reduces the viability of this strategy from the visual system's point of view. In particular, participants can only enumerate three groups of dots – two colors and the superset of all dots – in parallel (Halberda et al., 2006). So the "independently estimate then add" strategy suggested by (25b) predicts performance to decrease as the number of colors increases.

In contrast with this prediction, Lidz et al. (2011) observe identical performance regardless of the number of colors on screen. This does not appear to be a quirk of English: Tomaszewicz (2011) replicates this effect in Polish and Knowlton et al. (2018) observe the same effect in Cantonese (in both cases participants were asked to evaluate sentences containing a majority quantifier analogous to English *most*). These results suggest that participants are using a strategy more in line with a representation like (25c), where no reference is made to groups defined in part by predicate negation (e.g., the non-blue dots or the non-green frogs). For that reason, (25c) suggests a strategy that would remain comfortably under the 3-group limit regardless of the number of different colored groups presented on screen.

In the same vein, Knowlton et al. (2021a) offer more evidence against (25b). They report a series of experiments with adults and children that compared *most*-sentences to *more*-sentences (which have a meaning more in line with (25b)). Participants preferred to match sentences like *most of the dots are blue* with spatially-

intermixed pictures that made it easy to identify the superset of all dots, and sentences like *more of the dots are blue* with spatially-separated pictures where identifying the blue and the non-blue (i.e., the yellow) was easier. Moreover, when given the opportunity to create their own images – "make it true that more/most of the dots are blue" – participants created spatially-intermixed images for *most* but spatially-separated images for *more*. Likewise, when asked to recall properties of the blue and yellow dots, children who answered a *most* question were only able to remember the location of the blue dots, whereas those asked a truth-conditionally equivalent *more* question could recall the location of both blue and yellow.

Perhaps the most striking evidence comes from their final experiment. Adult participants were shown displays containing partially-intermixed blue and yellow dots, designed to make both a blue-yellow comparison strategy and a blue-total comparison strategy viable. But as mentioned above, a blue-yellow comparison strategy – like the one suggested by (25b) – is superior given only two colors. To see why, consider a display consisting of 20 blue dots and 10 yellow dots. Either strategy ultimately results in making the judgement that $20 > 10$. The negation strategy (in line with (25b)) does so directly, by representing the blue dots (20) and the non-blue dots (10). The subtraction strategy (in line with (25c)) does so indirectly, by comparing the number of blue dots (20) and the result of a subtraction $(30 - 20 = 10)$. The end result is the same, but because the Approximate Number System represents larger numerosities with more "noise" (see section 2.2.2 of Chapter 2), deriving 10 by subtracting 20 from 30 introduces more uncertainty to the system than representing 10 directly. The same is true for any two-color display: The

superset of all dots will always outnumber the yellow dots. So the direct blue-yellow comparison will always result in better performance than the analogous blue-total comparison.

Despite this, participants showed better performance when asked to evaluate *more*-sentences than when shown identical images and asked to evaluate truth-conditionally equivalent *most*-sentences. This is another case of participants eschewing a cognitively available and empirically superior strategy (direct blue-yellow comparison) in favor of a sub-optimal strategy (blue-total comparison). This result has a natural explanation if *most* has a meaning more like (25c) than (25b). That is, these results make sense if *most of the dots are blue* serves as an instruction to compare the cardinality of the blue dots against the cardinality of a subtraction as opposed to an instruction to compare the cardinality of the blue dots against the cardinality of the non-blue dots.

Finally, Hackl (2009) argues against a "more than half" specification, like (25a), using a novel "self-paced counting" task. Participants in one such experiment either heard sentences like *most of the dots are blue* or *more than half of the dots are blue* and were asked to verify the sentence by pressing "space" to reveal the color of masked dots, two or three at a time. Accuracies did not differ based on expression but reaction times did: Participants revealed dots at faster intervals when evaluating the *most*-sentence than when evaluating the truth-conditionally equivalent *more*-sentence. It remains unclear why reaction time differed in this task (see Talmina et al. (2017) for follow-up experiments using the "self-paced counting" paradigm and Steinert-Threlkeld et al. (2015) for follow-up experiments using a

paradigm closer to Pietroski et al. (2009) and Lidz et al. (2011)). But the finding that participants behave differently given truth-conditionally equivalent expressions is at least suggestive that *most* is not specified identically to *more than half*.[11]

Taken together, the evidence reviewed in this section points to *most* being specified in terms of subtraction, as in (25c) (abstracting away from the particular contributions of *frog* and *is green*). To repeat what this claim amounts to: The lexical item *most* has as its meaning a structured representation that combines with an internal predicate **I** and external predicate **E** and that makes reference to concepts like cardinality and subtraction, as in (27).

(27) $\lambda \mathbf{I}.\lambda \mathbf{E}.|\{x : \mathbf{I}(x) \; \& \; \mathbf{E}(x)\}| > |\{x : \mathbf{I}(x)\}| - |\{x : \mathbf{I}(x) \; \& \; \mathbf{E}(x)\}|$

As always, Fodor's challenge is present: It might be that *most* just points to an atomic MOST concept which entails (27). But this atomic concept would also need to entail the other possible specifications in (25), repeated here as (28).

(28) Most of the frogs are green

    a. $|\{x : Frog(x) \; \& \; Green(x)\}| > \frac{1}{2}|\{x : Frog(x)\}|$

    b. $|\{x : Frog(x) \; \& \; Green(x)\}| > |\{x : Frog(x) \; \& \; \neg Green(x)\}|$

    c. $|\{x : Frog(x) \; \& \; Green(x)\}| > |\{x : Frog(x)\}| - |\{x : Frog(x) \; \& \; Green(x)\}|$

    d. $OneToOnePlus(\{x : Frog(x) \; \& \; Green(x)\}, \{x : Frog(x) \; \& \; \neg Green(x)\})$

---

[11] Knowlton et al. (2021a) did not discuss "more than half", but data from their final experiment can be used to argue against a "more than half" specification. Namely, although *most* leads to worse performance than *more*, it does not lead to performance that is as bad as the division-based strategy suggested by (25a) would predict. Still, a more convincing case would require setting up the experiment such that *more than half*-sentences suggest a strategy that leads to superior performance and showing that, given *most*-sentences, participants nonetheless resort to the sub-optimal subtraction strategy suggested by (25c).

If most frogs are green, then the green frogs outnumber half of the frogs, they outnumber the non-green frogs, and they correspond one-to-one with the non-green frogs with at least one remainder. So if *most* has as its meaning an unstructured representation like MOST, one needs to say more to explain why participants seem to verify sentences like *most of the frogs are green* with a strategy transparent to (28c) even when superior alternatives corresponding to (28b) and (28d) are cognitively available.

A possible answer might be that (28c) is privileged in a way that other entailments are not. Certainly some entailments must be privileged, given that there are infinitely many of them (see Wellwood et al. (2015); Williams (2015) for discussion of this point in the verbal domain). To take an example, $2 + 2 = 4$ is entailed by *most of the frogs are green*, but it is unlikely that this particular mathematical fact is represented in the course of representing the expression. It might be that (28c) is different from $2 + 2 = 4$ in that it has special status as not only an entailment but a psychologically privileged one.

The question then becomes: What does it mean to be privileged? In the case of the lexical item *ewe* and the concept ANIMAL or the lexical item *bachelor* and the concept UNMARRIED, "privileged" might just mean "very strongly associated" (which would put them on a par with *superbowl* and SUNDAY, which also share a strong association in many speakers' minds). As discussed in section 1.3, for this reason it would be unsurprising to find that participants in an experiment token the concept UNMARRIED any time they hear the word *bachelor*. But there doesn't seem to be any sense in which (28c) is very strongly associated with *most of the frogs are*

green. So evidence that participants often seem to think (28c) when representing the relevant sentence is striking. And it suggests that "privileged" here more likely means "explicitly part of the representation."

Keeping this example in mind as a guide, we turn to the main topic of this dissertation: the mental representation of universal quantifiers. The initial hypothesis will be a psychologized version of the standard view, summarized in section 1.4. The alternative hypotheses considered here are presented in section 1.5.

1.4 Quantificational determiners as generalized quantifiers

Since Barwise and Cooper (1981), it has been standard to treat quantificational determiners like *every*, *most*, and *some* as instances of generalized quantifiers (see Westerståhl (2019) for a helpful review). On this view, a determiner like *every* expresses a two-place second-order relation, like (29), where $\mathbf{I}$ is the determiner's internal argument and $\mathbf{E}$ is its external argument.

(29) $[\![every]\!](\mathbf{I})(\mathbf{E}) = \text{TRUE} \iff \{x : \mathbf{I}(x)\} \subseteq \{x : \mathbf{E}(x)\}$

 $\approx \text{TRUE}$ *iff the set of I-things is a subset of or is identical to the set of E-things*

The specification in (29) is second-order because it relates two sets of individuals – one defined by $\mathbf{I}$ and one defined by $\mathbf{E}$ – as opposed to relating two individuals. And it's a relation because those two sets are individuated completely independently of one another and related by "$\subseteq$".

Generalized quantifiers were developed independently from concerns about natural language meaning (Mostowski, 1957). But applying the tools of General-

ized Quantifier Theory (GQT) to natural language has provided a useful framework for stating various generalizations (most prominently, those found in Barwise and Cooper (1981), Higginbotham and May (1981), and Keenan and Stavi (1986)). Sections 1.4.1 - 1.4.3 discuss some of the historical motivations for adopting GQT, and review some reasons for thinking they no longer hold. Section 1.4.4 recasts the GQT treatment of universal quantifiers (i.e., (29)) in overtly mentalistic terms so that it can serve as an initial hypothesis about their mental representation.

1.4.1 Expressive power and proportional quantifiers

Proportional quantifiers like *most* played a large role in the adoption of GQT. The reason is that while quantifiers like *every* and *some* can be easily modeled without invoking set theory, as in (30), *most* resists definition in first-order predicate logic (Rescher, 1962; Wiggins, 1980).

(30)   a. $[\![every]\!](\mathbf{I})(\mathbf{E}) = \text{TRUE} \iff \forall x[\mathbf{I}(x) \to \mathbf{E}(x)]$

          $\approx \text{TRUE}$ *iff for each thing, if it is I then it is E*

    b. $[\![some]\!](\mathbf{I})(\mathbf{E}) = \text{TRUE} \iff \exists x[\mathbf{I}(x) \ \& \ \mathbf{E}(x)]$

          $\approx \text{TRUE}$ *iff there exists some thing that is I and E*

That is, we cannot invent a first-order quantifier "$\mathcal{M}x$" that can be used to model *most* along the lines of "$\forall x$" and "$\exists x$" in (30). Replacing the general "$[\![most]\!](\mathbf{I})(\mathbf{E})$" with a specific example, we can consider two attempts to model the truth-conditions of *most frogs are green* in (31).

(31) Most frogs are green = $\text{TRUE} \iff$

a. $\mathcal{M}x[Frog(x) \rightarrow Green(x)]$

    $\approx$ *Most things are such that if they are a frog, they are green*

b. $\mathcal{M}x[Frog(x) \ \& \ Green(x)]$

    $\approx$ *Most things are such that they are a frog and they are green*

Both are insufficient. In the former case, (31a) is made incorrectly TRUE by a domain that includes a disproportionate number of non-frogs. Given 100 dogs, 1 green frog, and 10 blue frogs, *most frogs are green* is clearly FALSE, but (31a) is vacuously TRUE. And (31b) is made incorrectly FALSE in the same way. Given 100 dogs, 10 green frogs, and 1 blue frog, *most frogs are green* is TRUE, but (31b) is FALSE (most things aren't even frogs, let alone green ones).

One way out of this predicament would be to restrict the domain of quantification to just those things that satisfy the determiner's internal argument; just the frogs, in this case (Higginbotham and May, 1981). That is, determiners like *most* could be treated as devices for creating one-place restricted quantifiers (the main contention of Chapter 2 is that they should be, or at least that the universal quantifiers should be). Namely, although *most* cannot be specified with either (31a) or (31b), it can be specified as in (32).

(32) $\mathcal{M}x(Frog(x))[Green(x)]$

    $\approx$ *Most things that are frogs are such that they are green*

In (32), *frog* combines with "$\mathcal{M}x$" to form a restricted quantifier, "$\mathcal{M}x(Frog(x))$", which restricts the domain over which x ranges to just the (contextually relevant)

43

frogs.[12] Other quantifiers, like *every* and *some*, can be given analogous restricted specifications, as in (33).

(33)   a. $\forall x (Frog(x))[Green(x)]$

        $\approx$ *Each thing that is a frog is such that it is green*

       b. $\exists x (Frog(x))[Green(x)]$

        $\approx$ *Some thing that is a frog is such that it is green*

Another solution – the one that was pursued more fervently – is to turn to set theory. Instead of treating *every frog* as a one-place restricted quantifier, it can be thought of as a predicate of predicates whose extension is a set of sets (or, as Barwise and Cooper (1981) put it, a "family of sets"). For example, *every frog* is said to denote the set of sets whose members are all frogs (so if given a set, *every frog* returns TRUE if that set is in the mentioned set of sets, and FALSE if it is not). Then the determiner *every* can be thought of as a device for relating two sets, those provided by its internal and external arguments, as in (29), repeated here as (34a). This also allows for modeling proportional determiners like *most, more than half,* and *two thirds* along the same lines (e.g., in (34c)).

(34)   a. $[\![every]\!](\mathbf{I})(\mathbf{E}) = \text{TRUE} \iff \{x : \mathbf{I}(x)\} \subseteq \{x : \mathbf{E}(x)\}$

        $\approx$ TRUE *iff the set of I-things is a subset of or is identical to the set of*

        E-things

---

[12] The contribution of the experiments on *most* discussed in section 1.3.2 is to cash out how "$\mathcal{M}$" should be specified. That is, given that "things" is restricted to frogs, what does it mean for "most things" to be green? Does it mean that they correspond one-to-one-with-a-remainder to the non-green things? Or does it involve cardinality and subtraction? Put another way, the question of whether a quantifier should be specified in relational or restricted terms is a question at a higher level than the questions being asked in section 1.3.2.

b. $[\![some]\!](\mathbf{I})(\mathbf{E}) = \text{TRUE} \iff \{x : \mathbf{I}(x)\} \cap \{x : \mathbf{E}(x)\} \neq \emptyset$

$\approx$ TRUE *iff the intersection of the set of I-things and the set of E-things is non-empty*

c. $[\![most]\!](\mathbf{I})(\mathbf{E}) = \text{TRUE} \iff$

$|\{x : \mathbf{I}(x)\} \cap \{x : \mathbf{E}(x)\}| > |\{x : \mathbf{I}(x)\} - \{x : \mathbf{E}(x)\}|$

$\approx$ TRUE *iff the cardinality of the intersection of the set of I-things and the set of E-things is greater than the cardinality of the set of I-things minus the set of E-things*

In formal semantics, the set-theoretic approach has been preferred over the restricted quantification approach. Barwise and Cooper (1981) based their work exploring the linguistic consequences of GQT on ideas laid out in Montague (1973)'s influential paper, "The Proper Treatment of Quantification in Ordinary English". It's worth mentioning, in this context, that as far as Montague was concerned, the proper treatment of quantification was decidedly not a psychological treatment. In a history of semantics, Partee (2018) notes that Montague's theory was intended to be general enough to cover both natural and invented languages and that "Montague was surprised to learn that the linguists' notion of Universal Grammar was meant to capture all and only possible human languages; that struck him as parochial" (p.180). So, for followers of Montague, including Barwise and Cooper, generalized quantifiers had the virtue of being general.

To get a sense of how general, consider a few generalized quantifiers that have no natural language counterpart (at least among determiners), where A and B are

shorthand for $\{x : A(x)\}$ and $\{x : B(x)\}$:

(35) a. $Q(A)(B) = \textsc{true} \iff A = B$

  b. $Q(A)(B) = \textsc{true} \iff |A| > |B|$

  c. $Q(A)(B) = \textsc{true} \iff A \cap B \neq \emptyset \,\&\, |B| > 2$

  d. $Q(A)(B) = \textsc{true} \iff |A| - 1 = |B| \,\&\, |B| \text{ is prime}$

There are infinitely many others. In other words, whereas standard first-order predicate logic is insufficient to model determiners like *most*, GQT provides far more expressive power than is needed for modeling natural language determiners. This was by design, but whether it should be considered a desirable feature depends on the target of inquiry.

1.4.2 An analogy to transitive verbs

One initial point in favor of treating determiners under GQT was that it fit nicely with the largely relational conception of meaning that linguistics adopted from Frege (1879, 1893). An analogy to transitive verbs helps make the intuition clear: The verb *saw* in (36) can be thought of as expressing the SAW relation between two individuals, Kermit and Gritty. Likewise, the determiner *every* in (37) can be thought of as expressing the EVERY relation between two sets of individuals, the frogs and the green things.

(36) Kermit saw Gritty

```
                      S
                    /   \
                  NP     VP
                  |     /  \
                Kermit V    NP
                       |     |
                      saw  Gritty
```

(37) Every frog is green

```
                      S
                    /   \
                  DP     VP
                 /  \    / \
                D    NP is green
                |     |
              Every  frog
```
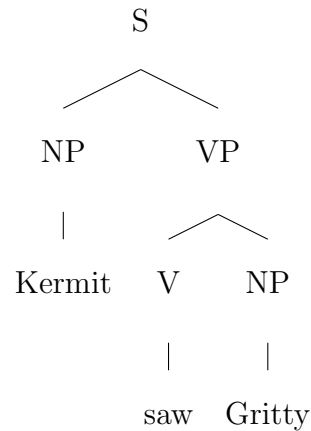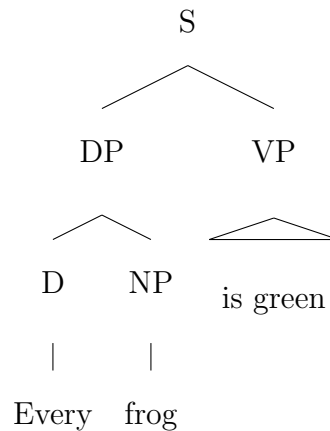
More formally, if proper nouns like *Kermit* and *Gritty* are names of entities (type e), then a transitive verb like *saw* can be thought of as a function from entities to a function from entities to truth values. That is, *saw* expresses the function of type $\langle e, \langle e, t \rangle \rangle$ in (38a).

(38) a. $[\![saw]\!] = \lambda x_e . \lambda y_e . \text{TRUE} \iff \text{SAW}(y, x)$

     b. $[\![every]\!] = \lambda X_{\langle e, t \rangle} . \lambda Y_{\langle e, t \rangle} . \text{TRUE} \iff \text{EVERY}(X, Y)$

Quantificational determiners can then be thought of as the same sort of functions, but with an extra level of abstraction, as in (38b). Namely, if common nouns

like *frog* and stative predicates like *is green* denote functions from entities to truth values (type $\langle e, t \rangle$), then a determiner like *every* can be thought of as a function from a type $\langle e, t \rangle$ function to a function from another type $\langle e, t \rangle$ function to truth values (i.e., *every* expresses the function of type $\langle \langle e, t \rangle, \langle \langle e, t \rangle, t \rangle \rangle$ in (38b), where EVERY is a generalized quantifier that relates X and Y via "⊆"). So *every* is just like *saw* except that the entities of type e are replaced with functions of type $\langle e, t \rangle$ and that the order of the arguments is switched on the right side of the lambdas.[13]

With the rise of event semantics though, this analogy no longer goes through. In particular, on modern "Neo-Davidsonian" views (alluded to in section 1.2.1), transitive verbs like *saw* are not thought of as expressing relations between individuals. And on some variants they are not thought of as expressing relations at all. There are different approaches, but a common theme is separating NP arguments from verbs (see Chapter 9 of Williams (2015) for discussion of arguments in favor of such separation and Williams (2021) for a helpful review of event semantics).

Kratzer (1996), for example, advocates severing the external argument of the verb (e.g., *Kermit* in (39)) and instead using a distinct functional head (v) to introduce it as an argument of the silent operator AG (short for AGENT). So (39) would be analyzed as (40).

---

[13] The switch is a result of *every* being said to correspond to "⊆" not "⊇". In keeping with the convention that the external argument of a relation appears first, as in *Kermit saw Gritty* being regimented as SAW(K, G), *every frog is green* should really be written EVERY(G, F). But then EVERY would correspond to "⊇", not "⊆", as is traditionally assumed.

(39) Kermit saw Gritty

vP

NP $\bar{v}$

|

Kermit v VP

|

AG   V   NP

| |

saw   Gritty
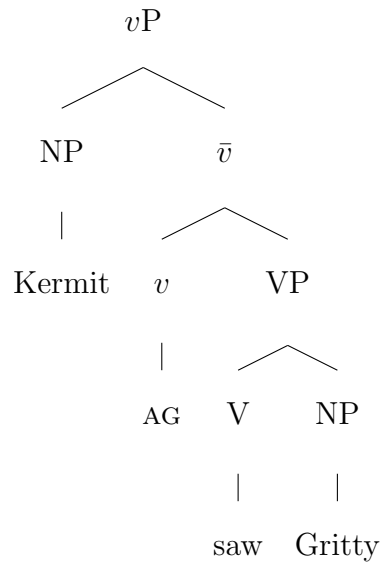
(40) $\exists e[\text{AGENT}(e, Kermit)\ \&\ \text{SAW}(e, Gritty)]$

$\approx$ *there exists an event s.t. Kermit is its agent and it was a seeing of Gritty*

Instead of being a function from entities to a function from entities to truth values (type $\langle e, \langle e, t \rangle \rangle$), a transitive verb like *saw* is treated as a function from entities to a function from events (type v) to truth values (type $\langle e, \langle v, t \rangle \rangle$), as in (41) (which combines with *Gritty* through function application and with $\text{AGENT}(e, Kermit)$ through conjunction to yield (40)).

(41) $[\![saw]\!] = \lambda x_e.\lambda e_v.\text{TRUE} \iff \text{SAW}(e, x)$

Given (41), we can no longer say that the verb SAW expresses a relation between individuals. Instead, it expresses a relation between an individual and an event. Of course, any event that is a seeing involves a relation between a seer and a seen, so "$\text{SAW}(e, x)$" entails these relations. Still, the verb itself does not denote a relation between two entities.

Other approaches go further. Schein (1993) and Krifka (1992), for example, offer evidence for severing both the internal and the external argument from the verb. One possibility for spelling out this view is saying that NP arguments come marked with thematic features (e.g., $Kermit_{\text{AGENT}}$ and $Gritty_{\text{PATIENT}}$) and that these features determine the thematic predicate with which they combine (Hornstein, 2002). Assuming complete separation between a verb and its arguments means that a sentence like *Kermit saw Gritty* would be analyzed as (42).[14]

(42) $\exists e[\text{AGENT}(e, Kermit) \ \& \ \text{PATIENT}(Gritty) \ \& \ \text{SAW}(e)]$

$\approx$ *there exists an event s.t. Kermit is its agent and Gritty is its patient and it was a seeing*

On such a view, a transitive verb like *saw* doesn't take any NP arguments and can't be said to express a relation at all. Instead, it expresses a monadic predicate of events, as in (43).

(43) $[\![saw]\!] = \lambda e_v.\text{TRUE} \iff \text{SAW}(e)$

Again, this is not to deny that any event of seeing involves a relation between seer and seen. We can say that "$\text{SAW}(e)$" entails such a relation without saying that the verb itself denotes a relation. Any event that is a seeing also involves a location in space and a time, but this hardly motivates treating *saw* as a four-place relation between a seer, a seen, a location, and a time (see Williams (2015), in particular

---

[14] Given this view, a rule of interpretation rooted in predicate conjunction seems more natural for combining the verbal predicate and its thematic predicates than a rule of interpretation rooted in function application (Pietroski 2005; LaTerza 2014). Though both are in principle possible.

Chapter 4). It is enough to notice that these roles are entailed, without the further commitment of explicitly encoding them in the representation.

On modern approaches, then, transitive verbs either don't express relations between individuals or don't express relations at all. To the extent that these approaches are on the right track, the analogy between quantificational determiners as relations and transitive verbs as relations no longer seems as compelling. This is not itself an argument against GQT; just an argument against using the analogy to transitive verbs to motivate it. But if it turns out that relationality in general is less prevalent in natural language than has been assumed (e.g., as Pietroski (2018) argues), then it may be preferable to consider non-relational treatments of quantification as well.

1.4.3   Mismatch between grammatical and logical form

Another point often used to motivate GQT is, as Barwise and Cooper (1981) put it, "the notorious mismatch between the syntax of noun phrases in a natural language like English and their usual representations in traditional predicate logic" (p.164). For example, consider the syntax of *every frog is green* and its predicate logic treatment in (44).

(44)
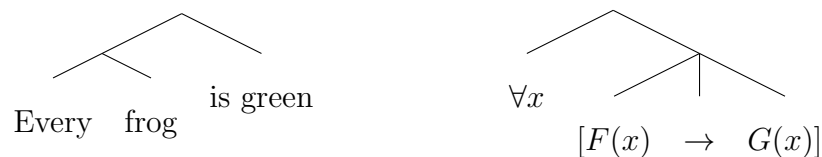
The mismatch Barwise and Cooper have in mind is twofold. One issue is that even if

"$\forall x$", "$F(x)$", and "$G(x)$" can each be thought of as corresponding to a grammatical

constituent (*each*, *frog*, and *is green*), the connective "$\rightarrow$" cannot. There is no node

in the tree on the left that corresponds to "$\rightarrow$". A second issue is that no aspect

of the predicate logic treatment corresponds to the grammatical constituent *every*

frog. Instead, "$F(x)$" and "$G(x)$" play the same logical role: There is no sense in

which "$\forall x$" is more closely related to "$F(x)$" than it is to "$G(x)$".

At least with respect to the first issue, GQT does a better job of reflecting the

syntax, as in (45).

(45) Generalized Quantifier Theory

$$
\begin{array}{c}
\text{S} \\
\diagup\diagdown \\
\text{DP} \qquad\qquad \text{VP} \\
\diagup\diagdown \qquad\qquad \diagup\diagup\diagdown \\
\text{D} \qquad \text{NP} \qquad \text{is green} \\
| \qquad\quad | \\
\text{Every} \quad \text{frog} \qquad \{x : G(x)\} \\
\\
\subseteq \qquad \{x : F(x)\}
\end{array}
$$
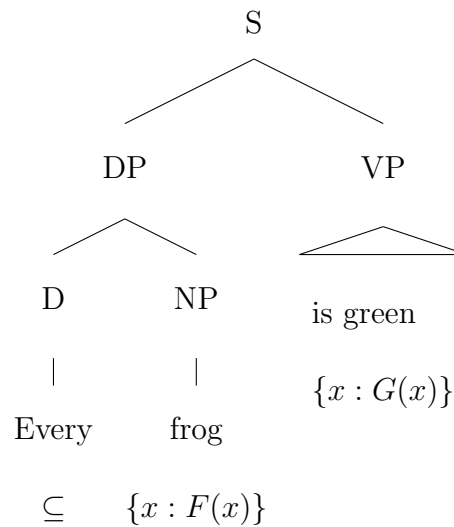
That is, there is no logical connective in the GQT treatment in (45), so the prob-

lematic "$\rightarrow$" has been dealt with.

The second issue however – that *every frog* is not a constituent in the predicate

logic treatment – does not seem to be solved by invoking the relational notion "$\subseteq$",

since "$\subseteq (\{x : F(x)\})$," is no more a constituent than "$\forall x[F(x)$". But given that

every frog is taken to denote a function from sets of individuals to truth values, it can

be thought of as a family of sets to which the set denoted by *is green* is related. As

Barwise and Cooper put it, on GQT "the sentences [{*some person/every man/most babies*} *sneeze*] will be true just in case the set of sneezers ... contains some person, every man, or most babies" (p.165). In this sense, the quantificational DP can be thought of as a constituent in the logical form as well.

Of course, GQT is not the only way out of the apparent mismatch. Invoking restricted quantification as in (46) has the effect of (i) removing the problematic "→" and (ii) treating the internal argument (*frog*) as logically more closely related to *every* than the external argument (*is green*).

(46) Restricted quantification
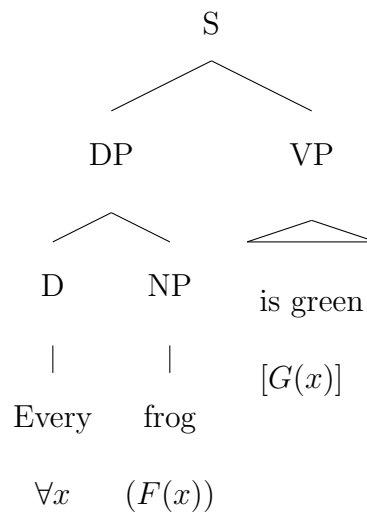
The second point requires some elaboration. Restricted quantification was alluded to in section 1.4.1 and will be discussed at length in Chapter 2, but in short, its defining feature is that the two arguments of the determiner play different logical roles. In this case, F initially restricts the values of x, whereas G functions more like a normal predicate, returning TRUE just in case the value taken on by x is green. So the symbols in the logical form corresponding to the DP – "$\forall x(F(x))$" – can be thought of as a prefix to the sentence specified by the external argument: "$G(x)$".

In this way, *every* and *frog* are more closely related logically than *every* and *is green*.

Having said that, the point may be moot. Assuming categorematic meanings for determiners, like "$\lambda \mathbf{I}.\lambda \mathbf{E}.\text{EVERY}(\mathbf{I}, \mathbf{E})$", what appears "on the right side" of the lambdas doesn't matter as far as the syntax is concerned. If the predicate logic or GQT treatment is viewed not as a logical form, but as a specification of EVERY, then the details of this specification have no relation whatsoever to syntax. That is, cashing out "$\text{EVERY}(\mathbf{I}, \mathbf{E})$" in (47) as "$\mathbf{I} \subseteq \mathbf{E}$" is no better or worse a reflection of grammatical form than cashing it out as "$\forall x[\mathbf{I}x \rightarrow \mathbf{E}x]$" or as "$\forall x(\mathbf{I}x)[\mathbf{E}x]$".

(47)

```
                              S
                    ┌─────────────────────┐
                   DP                      VP
              ┌──────────┐            ┌─────────┐
              D          NP            is green
              |          |           λy.Green(y)
            Every       frog
```

$\lambda \mathbf{I}.\lambda \mathbf{E}.\text{EVERY}(\mathbf{I}, \mathbf{E}) \quad \lambda x.Frog(x)$
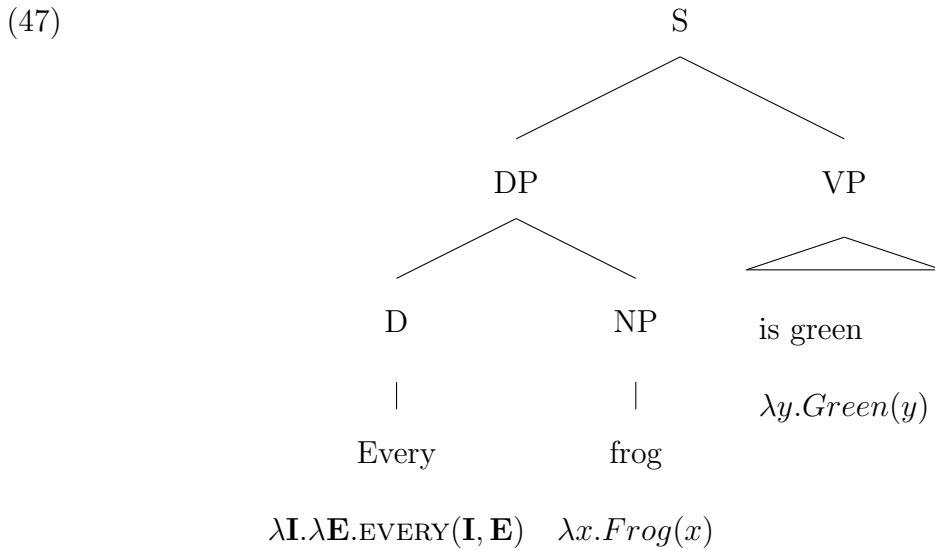
This is true regardless of whether the details of how to specify EVERY are taken to be a theorist's choice about how to specify a function in extension or a psychological hypothesis about the format of a mental representation.

So, from the point of view of matching grammatical and logical form, GQT and restricted quantification are on a par.

1.4.4 Psychologizing Generalized Quantifier Theory

Given its popularity, it seems reasonable to use the GQT treatment of universal quantifiers, repeated in (48), as our initial hypothesis about how those quantificational determiners are mentally represented.

(48) $[\![each/every/all]\!](\mathbf{I})(\mathbf{E}) = \text{TRUE} \iff \{x : \mathbf{I}(x)\} \subseteq \{x : \mathbf{E}(x)\}$

$\approx \text{TRUE}$ *iff the set of I-things is a subset of or is identical to the set of E-things*

The two main features of this view are that determiners are devices for relating two things (e.g., as opposed to being devices for creating restricted quantifiers) and that the relation is second-order (e.g., between sets of individuals instead of between individuals).

In principle, the GQT treatment might be an accurate description of how speakers mentally represent quantificational determiners along one of these dimensions but not the other. It might be that speakers represent *every frog* as a second-order restricted quantifier, for example. But set theory and second-order relations go hand-in-glove. So the use of set theory in (48) makes these two notions – relational vs. restricted and first-order vs. second-order – hard to disentangle. This is not a problem if the goal is to articulate a specification of truth in a model, since the choice of how to write (48) is merely a notational one. But when the target of inquiry is a lexical item's representational format, divorcing the notion of relationality from the notion of second-order quantification becomes important.

To that end, this dissertation will formulate GQT in terms of plural logic

instead of set theory. Intuitively, this amounts to replacing "the set of **I**-things" with "the **I**s" and "the set of **E**-things" with "the **E**s". Not only will this allow us to more naturally distinguish along both dimensions of interest, it will also allow us to remain agnostic about how second-order representations should be understood. Chapter 3 discusses this point in more detail, but in short: The plural notion "the **I**s" can be taken to implicate a plural entity like a set or it can be taken to implicate a plurality of individuals that is not itself an entity. Remaining agnostic on this point is preferable. While it is widely agreed in cognitive science (and supported by intuition) that human minds can form some kind of group representation (see section 3.2 of Chapter 3), there is no reason to think these representations have all of the formal properties that would make them sets.

Formally, instead of treating the satisfiers of a predicate **P** as constituting the set of **P**-things, as in (49a), we can group them under a second-order variable G (for "group"), as in (49b). This second-order variable in (49b) could be understood as ranging over sets – in which case it would be a notational variant of (49a) – or as ranging over individuals but allowing more than one value to be assigned to a single variable per assignment (Boolos, 1984).

(49)   a. $\{x : \mathbf{P}x\}$ ("the set of **P**-things")

        $\approx$ *the set of things that meet the $\boldsymbol{P}$ condition*

    b. $\iota G(\forall x[Gx \equiv \mathbf{P}x])$ ("the **P**s")

        $\approx$ *the one or more things$_G$ s.t. for each thing$_x$, it$_x$ is one of them$_G$ iff it$_x$*

        meets the $\boldsymbol{P}$ condition

This change might be relevant for future work aimed at distinguishing hypotheses that differ with respect to whether they rely on a conjunctive description, like (50a), or intersecting two groups, like (50b).

(50) a. $\{x : \mathbf{I}x \ \& \ \mathbf{E}x\}$

  b. $\{x : \mathbf{I}x\} \cap \{x : \mathbf{E}x\}$

In one sense, these two seem different. But given standard set theory, both (50a) and (50b) are names for a single set. The set of things that are frogs and green, for example, just is the intersection of the set of frogs and the set of green things. On the other hand, the plural logic versions in (51a) and (51b) make the distinction clearer: The former only has quantification into one second-order variable position, "$\iota G(...G...)$", whereas the latter has quantification into two such positions, "$\iota G(...G...)$" and "$\iota G'(...G'...)$". Put another way, whereas (51a) implicates only one group, (51b) clearly implicates two groups.

(51) a. $\iota G(\forall x[Gx \equiv \mathbf{I}x \ \& \ \mathbf{E}x])$

   $\approx$ *the one or more things$_G$ s.t. for each thing$_x$, it$_x$ is one of them$_G$ iff it$_x$ meets the **I** condition and the **E** condition*

  b. $\iota G(\forall x[Gx \equiv \mathbf{I}x]) \cap \iota G'(\forall x[G'x \equiv \mathbf{E}x])$

   $\approx$ *the one or more things$_G$ s.t. for each thing$_x$, it$_x$ is one of them$_G$ iff it$_x$ meets the **I** condition intersected with the one or more things$_{G'}$ s.t. for each thing$_x$, it$_x$ is one of them$_{G'}$ iff it$_x$ meets the **E** condition*

This particular distinction is not explored in this dissertation. But that the plural

logic formulation can more easily highlight the difference is a point in its favor, at least for the purposes of theorizing about mental representations.

To be sure, the main claims of this dissertation could be recast in set theoretic terms. The simplest way to do so would be to interpret quantification into second-order variable positions – like "$\iota G(...G...)$" – as quantification over sets (i.e., to read it as "the set G s.t. ..."). But putting things in the slightly more cumbersome terms of plural logic has at least a few virtues. It allows us to more easily distinguish the dimensions of interest, prevents us from making an extra ontological commitment, and reminds us that the symbols represent mentalistic hypotheses.

One quick note about notation is in order. Brackets – $(...), [...], \{...\}$ – will be used interchangeably and alternated or omitted for readability. So the difference between, for example, (52a) and (52b) is merely a notational difference and the two specifications do not correspond to distinct psychological hypotheses.

(52)   a. $Qx[\mathbf{I}(x)]\{\mathbf{E}(x)\}$

          $\approx Q$ *thing that is an* $\mathbf{I}$ *is s.t. it is an* $\mathbf{E}$

    b. $Qx(\mathbf{I}x)(\mathbf{E}x)$

          $\approx Q$ *thing that is an* $\mathbf{I}$ *is s.t. it is an* $\mathbf{E}$

With all of this in place, we can specify a psychologized version of the GQT treatment of universal quantifiers in (53).

(53) $\iota G(\forall x[Gx \equiv \mathbf{I}x])\{\iota G'(\forall y[G'y \equiv \mathbf{E}y])[G \subseteq G']\}$ (SO relational, 2G)

    $\approx$ *the one or more things$_G$*

    (s.t. for each thing$_x$, it$_x$ is one of them$_G$ iff it$_x$ meets the $\mathbf{I}$ *condition)*

are s.t. the one or more things$_{G'}$

(s.t. for each thing$_x$, it$_x$ is one of them$_{G'}$ iff it$_x$ meets the **E** condition)

are s.t. they$_G$ are among them$_{G'}$

This makes explicit the central claims of GQT: (53) is a second-order relation between two groups (SO relational, 2G). We can then formulate alternative hypotheses that differ wither respect to (i) whether the representation is first-order or second-order, (ii) whether the representation is a relation, and (iii) how many groups are implicated. This is taken up in section 1.5.

1.5   Ungeneralizing the universal quantifiers

As noted above, one way for a specification to differ from (53) is for it to be restricted instead of relational. Consider (54). Instead of specifying the quantificational content by relating two independent groups, G and G', (54) does so by first restricting the values of G to just the things that meet the **I** condition, then noting that "the **I**s" (i.e., the one or more things$_G$) also meet the **E** condition.

(54) $\iota G(\forall x[Gx \equiv \mathbf{I}x])[\mathbf{E}(G)]$ (SO restricted, 1G)

$\approx$ the one or more things$_G$

(s.t. for each thing$_x$, it$_x$ is one of them$_G$ iff it$_x$ meets the **I** condition)

are s.t. they$_G$ meet the **E** condition

This way of specifying restricted quantification is still second-order, because it implicates a group G (see Chapter 3 for discussion of the distinction between first-order and second-order quantification).

Of course, (54) also differs from (53) in that it makes reference only to a single group. We could instead imagine a version more similar to (53) that makes reference to two groups but is nonetheless restricted in the relevant sense. For example, consider (55).

(55) $\iota G(\forall x[Gx \equiv \mathbf{I}x])\{\iota G'(\forall y[G'y \equiv Gy \ \& \ \mathbf{E}y])[G = G']\}$ (SO restricted, 2G)

$\approx$ *the one or more things$_G$*

*(s.t. for each thing$_x$, it$_x$ is one of them$_G$ iff it$_x$ meets the **I** condition)*

are s.t. the one or more things$_{G'}$

*(s.t. for each thing$_x$, it$_x$ is one of them$_{G'}$ iff it$_x$ is one of them$_G$ and it$_x$ meets the **E** condition)*

are s.t. they$_G$ are identical to them$_{G'}$

This specification is restricted because the values of the second group, G', are restricted by the values of the first group, G. So "$\iota G(...G...)$" and "$\iota G'(...G'...)$" are not independent selections from some domain; only "$\iota G(...G...)$" is. The quantificational content is specified in terms of identity: "$[G = G']$". To the extent that identity is a relation, perhaps we should distinguish restricted vs. unrestricted from relational vs. non-relational. But in the sense relevant for this dissertation, the representation in (55) is restricted, not relational. It differs from the restricted (54) in that it implicates two groups instead of only one.

Lastly, we can consider a restricted specification – (56) – that implicates zero groups, and consequently is completely first-order.

(56) $\forall x(\mathbf{I}x)[\mathbf{E}x]$ (FO restricted, 0G)

   $\approx$ *each thing$_x$ that meets the $\mathbf{I}$ condition is s.t. it$_x$ meets the $\mathbf{E}$ condition*

This specification will be familiar from section 1.4.1. It is restricted because the values of x limited just to things that meet the $\mathbf{I}$ condition. This makes it importantly different from the more familiar unrestricted (57).

(57) $\forall x[\mathbf{I}x \to \mathbf{E}x]$ (FO unrestricted, 0G)

   $\approx$ *each thing$_x$ is s.t. if it$_x$ meets the $\mathbf{I}$ condition, it$_x$ meets the $\mathbf{E}$ condition*

1.6   Chapter summary and dissertation overview

Semantic hypotheses are often not treated as claims about what speakers mentally represent. But speakers do mentally represent linguistic meanings, and those representations must be encoded in some format. This chapter discussed how to treat formal semantic claims as psychological hypotheses, and reviewed some examples from linguistics and psycholinguistics.

Probing representational format can be difficult. The evidential standards are high and the alternative hypotheses are ubiquitous. But compelling cases can be made for thinking a given lexical item has one particular "psycho-logical form" over another. Hopefully this dissertation provides one such case.

The main question will be how the universal quantifiers are represented. Supposing a quantificational determiner like *every* in some sense relates two predicates, as in (58), we can ask: How should we specify what occurs on the right side of the lambda expressions?

(58) $[\![every]\!] = \lambda\mathbf{I}.\lambda\mathbf{E}.\text{EVERY}(\mathbf{I}, \mathbf{E})$

This chapter introduced a few hypotheses, given again below:

(59) $\iota G(\forall x[Gx \equiv \mathbf{I}x])\{\iota G'(\forall y[G'y \equiv \mathbf{E}y])[G \subseteq G']\}$ $\qquad$ =(53) SO relational, 2G

(60) $\iota G(\forall x[Gx \equiv \mathbf{I}x])\{\iota G'(\forall y[G'y \equiv Gy \ \& \ \mathbf{E}y])[G = G']\}$ =(55) SO restricted, 2G

(61) $\iota G(\forall x[Gx \equiv \mathbf{I}x])[\mathbf{E}(G)]$ $\qquad\qquad$ =(54) SO restricted, 1G

(62) $\forall x(\mathbf{I}x)[\mathbf{E}x]$ $\qquad\qquad$ =(56) FO restricted, 0G

As a starting point, we will treat the standard semantic account of quantifiers – Generalized Quantifier Theory – as a psychological hypothesis about their mental representation. To be sure, this account was not originally intended to be understood as a claim about mental representation. Nonetheless, it can easily be recast in mentalistic terms. This "psychologized" version of GQT in (59) holds that sentences like *every frog is green* are mentally represented as two-place second-order relations. This dissertation puts that hypothesis to the test and ultimately argues against it.

Chapter 2 considers the "psycho-logical" distinction between relational and restricted quantification (i.e., (59) vs. (60) - (62)). Evidence in support of the latter view comes in large part from a series of psycholinguistic experiments in which participants are asked to evaluate sentences like *every big circle is blue*. These experiments demonstrate that participants mentally represent the group picked out by the quantifier's internal argument (the big circles, in this example) but not the group picked out by its external argument (the blue things, in this example). Given that the visual system supports encoding multiple groups in parallel, this strategy

would be surprising if participants understood the sentence to express a relation between two groups.

The proposed non-relational alternative also receives support from more traditional linguistic evidence. In particular, it offers a satisfying explanation of the semantic universal "conservativity." This robust cross-linguistic generalization seems to hold of all natural language determiners: their internal argument can be duplicated within their external argument without a change in truth-conditions. For example, if *every frog is green* is TRUE, then *every frog is frog that is green* will likewise be TRUE. Plenty of relations fail to be "conservative" in this sense (e.g., identity; we can imagine a situation in which *the frogs are identical to the green things* is FALSE but *the frogs are identical to the green frogs* is TRUE). So if determiners express relations, more needs to be said to explain why they are only able to express certain relations. Alternatively, if determiners are tools for combining with an argument to form one-place restricted quantifiers, the "conservativity" generalization follows as a logical consequence.

Chapter 3 considers the "psycho-logical" distinction between first-order (i.e., individual-implicating; (62)) and second-order (i.e., group-implicating; (59) - (61)) quantification. Again, evidence comes in part from a series of psycholinguistic experiments in which participants are asked to evaluate quantificational statements. Though they always seem to represent the extension of the internal argument, they do so in different ways depending on the particular quantifier used. Sentences in which universal quantification is indicated by *every* or *all* encourage participants to rely on psychological systems for representing groups (ensemble representations).

In the very same situations though, sentences in which universal quantification is indicated by *each* lead participants to instead rely on their psychological system for representing individuals (object-file representations).

This distinction between first-order and second-order quantification is also argued to explain certain purely linguistic phenomena. First, it can help make sense of why *each* always gives rise to distributive interpretations in which the predicate applies to each individual, even in cases where *every* and *all* can give rise to a collective interpretation in which the predicate applies to the group as a whole. To be sure, this is not to say that facts pertaining to distributivity can be wholly explained by the first-order/second-order distinction (notions of distributive predicates and a distributivity operator are still required).

Second, the first-order/second-order distinction offers a new way to view subtle judgments about the ability of *each* and *every* to allow for certain kinds of generic interpretations. For example, *gravity acts on every object* can naturally be understood as a general claim that projects beyond the local domain, whereas *gravity acts on each object* feels odd, as if we were talking about each individual object in the universe, rather than stating a principle. This is argued to result from properties of the cognitive systems (ensemble and object-file representations) that first-order and second-order representations naturally interact with. If right, these sorts of generic interpretations have been misdiagnosed as purely linguistic facts, and actually have explanations that lie outside of grammar.

If the proposed representations for *each*, *every*, and *all* are on the right track, they suggest that knowing the meaning of these quantifiers requires associating,

64

with each pronunciation, a representation specified in a particular format (as opposed to, for example, associating them with universal truth-conditions). This raises an acquisition question: What information in their input would lead a learner to associate *each* with a first-order restricted universal concept and *every* with a logically-equivalent second-order restricted universal concept?

To begin to answer this question, Chapter 4 provides an initial sketch of how learners might acquire *each* and *every* (*all* is left for future work). In particular, an examination of corpora of child-ambient speech reveals relatively low-level distributional differences between uses of *each* and *every*. Most notably, parents often quantify over times with *every* (e.g., *every time we go to the store, you cry!*) and individuals with *each* (e.g., *pour milk into each cup*). This is argued to result from the genericity asymmetry discussed in Chapter 3 and an outline for how learners may be able to use these cues to select the correct representation to pair with *every* is presented.

This case study also informs bigger picture questions concerning how meanings are related to non-linguistic cognition. Much like phonological representations can be thought of as providing instructions to the motor-planning system, the present results support the theoretical idea that meaning representations can be thought of as providing instructions to conceptual systems. Methodologically, the results lend support to the linking hypothesis – the Interface Transparency Thesis – discussed in section 1.3.1. And more generally, to the extent that this sort of investigation is fruitful, it supports the idea that formal approaches to linguistic meaning can benefit from embracing, instead of avoiding, details of human psychology.

Chapter 2: Relational vs. restricted quantification

Quantificational determiners are standardly thought to express second-order relations (see section 1.4 of the previous chapter). This chapter takes on the relational component of this standard view, contrasting it with the alternative that determiners are devices for creating restricted quantifiers. It argues that restricted quantification is preferred both on semantic and psychosemantic grounds.

Section 2.1 discusses the logical distinction between relational and restricted quantification, which comes down to the logical role played by the two arguments. Section 2.2 presents a series of seven experiments aimed at determining which of the two arguments participants explicitly represent. In particular, the experiments use cardinality knowledge as a proxy for argument representation, so the psychological system for representing cardinality – the Approximate Number System – is introduced before discussing the results. To preview, when verifying sentences like *every big circle is blue*, participants only seem to represent the internal argument (big circle), not the external argument (blue things), and not the group named by the conjunction of both arguments (big blue circles). These results points toward quantificational determiners like *every* having restricted meanings. Section 2.3 then discusses the universal generalization that all determiners in natural language

have "conservative" meanings. Whereas relational quantification creates difficulties for explaining this phenomenon, restricted quantification has "conservativity" as a logical consequence.

2.1   The logical distinction

As discussed in Chapter 1, sentences like *every frog is green* and *Kermit saw every frog* have the structure in (63), where *every* combines with two arguments, an internal argument **I** (e.g., *frog*) and an external argument **E** (e.g., *t is green* or *Kermit saw t*, assuming quantifier raising).

(63)

```
                    S
                  /   \
               DP      VP
              /  \      |
            D    NP     E
            |     |
          Every   I
```
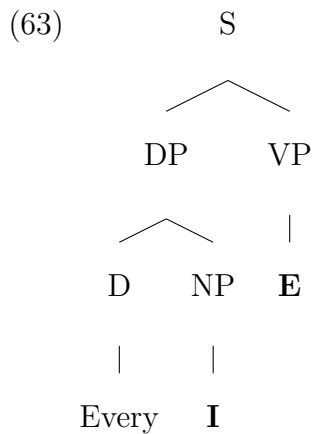
Both arguments name predicates, and the determiner, in some sense, contributes details about how the two predicates are related. The question at issue here is how to specify that relation between predicates.

The standard view, discussed in section 1.4 of the preceding chapter, is that determiners like *every* are tools for expressing genuine relations between the two predicates (i.e., they are special cases of generalized quantifiers). This raises the

question of why only a small fraction of all possible relations get lexicalized (an issue raised last chapter in section 1.4.1 and taken up again in this chapter in section 2.3). The alternative, older view, is that determiners are tools for creating restricted quantifiers. That is, the condition supplied by the internal predicate restricts the domain in which the expression is evaluated and the condition supplied by the external predicate adds a further condition that some quantity of the restricted domain must meet (e.g., *all* of them, *some* of them, *none* of them, *most* of them, *half* of them).

At least since Barwise and Cooper (1981), the relational view has been more popular in semantics. On this view, quantificational determiners like *every* express relations between groups.[1] Formally, this idea is often captured in set-theoretic terms, as in (64), leading to the meaning for *every frog is green* in (65).

(64) $[\![\mathbf{Every}]\!] = \lambda\mathbf{I}.\lambda\mathbf{E}.\{x : \mathbf{I}x\} \subseteq \{x : \mathbf{E}x\}$

    $\approx$ *the set of **I**-things is a subset of the set of **E**-things*

(65) $\{x : \mathbf{frog}(x)\} \subseteq \{x : \mathbf{green}(x)\}$

    $\approx$ *the set of frogs is a subset of the set of green things*

Following the notational conventions established in Chapter 1, we can state the same idea with the representation of universal quantification in (66).

---

[1] As noted in Chapter 1, the use of "group" instead of "set" is to remain agnostic between two ways of interpreting second-order variables (like G and G' in (66)). The relevant distinction is discussed at length in Chapter 3. In short, "group" can be understood to mean "some sort of plural entity, like a set" or "a way of grouping the relevant predicate by allowing more than one value to be assigned to a single variable per assignment."

(66) $\lambda \mathbf{I}.\lambda \mathbf{E}.\iota G(\forall x[Gx \equiv \mathbf{I}x])\{\iota G'(\forall y[G'y \equiv \mathbf{E}y])[G \subseteq G']\}$

$\approx$ *the things$_G$ that are $\mathbf{I}$ are among the things$_{G'}$ that are $\mathbf{E}$*

This representation explicitly groups and relates the predicates named by the two arguments. The things that satisfy the $\mathbf{I}$ condition are grouped under the variable G (with "$\iota G(\forall x[Gx \equiv \mathbf{I}x])$"). The things that satisfy the $\mathbf{E}$ condition are grouped under the variable G' (with "$\iota G'(\forall y[G'y \equiv \mathbf{E}y])$"). And the quantificational content is specified by requiring that the Gs be among/be a subset of the G's ("$[G \subseteq G']$"). In other words, the representation in (66) specifies the relation between the predicates supplied by the internal and external arguments in genuinely relational terms. Both arguments are treated as independent groups and the determiner itself is treated as a device for relating those groups in a particular way.

But the relation between the predicates supplied by $\mathbf{I}$ and $\mathbf{E}$ can also be specified without appealing to a relational notion like "$\subseteq$". Instead, as Higginbotham and May (1981) put it, the internal argument of the determiner can be thought of as "restrict[ing] the domain over which the variable bound by [the quantifier] ranges" (p.54). As Lepore and Ludwig (2007) similarly say: The internal argument of *all* in the sentence *all men are mortal* "functions as if it were a variable restricted to taking on as values only men" whereas in *all things are mortal*, the internal argument *things* "functions as a variable which can take on anything as a value" (p.61).

There are many ways of cashing out this restricted quantification idea. In set-theoretic terms, the difference between relational and restricted quantification can be thought of as the difference between making two independent selections from a

universe of evaluation and making an initial selection that limits the universe from which the second selection is made. To illustrate, consider two ways of specifying *every frog is green* in (67a) and (67b) (where the "hook" notation "↾ $\underline{u}$" signifies relativization to a particular universe of evaluation (Westerståhl, 2019)).

(67)   a. EVERY$_x$($\{x : \mathbf{frog}(x)\}, \{x : \mathbf{green}(x)\}$) ↾ $\underline{universe}$

        ≈ *Relative to the universe, everything in the set of frogs is also in the set of green things*

        b. EVERY$_x$($\{x : \mathbf{green}(x)\}$) ↾ $\underline{\{x : \mathbf{frog}(x)\}}$

        ≈ *Relative to the set of frogs, everything is in the set of green things*

In (67a), *every* supplies a two-place relation, which is evaluated with respect to the entire universe (which may be contextually restricted). In (67b) though, *every* is a monadic quantifier akin to *everything*, but evaluated only with respect to the set of frogs (plus any relevant contextual restriction). In that way, the internal argument – *frog* – restricts the values over which x can range. Both notions involve two sets, but only on the relational view is the second set (the set of green things) independent of the first (the set of frogs).

Moving away from set theory, the same intuition about restricted quantification can be stated in terms of quantifying over sequences of assignments of values to variables (Tarski, 1956). Suppose a sequence satisfies (68) iff it assigns x something green.

(68) *green*(x)

     ≈ *x is green*

(69) a. $\langle x \leftarrow Kermit; y \leftarrow Mr.Toad; z \leftarrow Gritty \rangle$ satisfies (68)

   b. $\langle x \leftarrow Gritty; y \leftarrow Mr.Toad; z \leftarrow Kermit \rangle$ doesn't satisfy (68)

To keep things simple, imagine a domain that only includes Kermit (a green frog), Mr. Toad (another green frog), and Gritty (an orange creature). The sequence in (69a) then satisfies (68), because it assigns something green to x; whereas the sequence in (69b) does not, because it assigns something orange to x.

We can form a new sentence, (70), by adding the prefix "$\forall x$" to (68). This sentence is satisfied if everything in the domain is green. Formally: (70) is satisfied by a sequence iff each x-variant of that sequence – i.e., each sequence that is just like it except with regard to what it assigns to x – assigns x something green.

(70) $\forall x[green(x)]$

   $\approx$ *everything is green*

(71) a. $\langle x \leftarrow Kermit; y \leftarrow Mr.Toad; z \leftarrow Gritty \rangle$ doesn't satisfy (70)

   b. $\langle x \leftarrow Mr.Toad; y \leftarrow Mr.Toad; z \leftarrow Gritty \rangle$ an x-variant of (71a)

   c. $\langle x \leftarrow Gritty; y \leftarrow Mr.Toad; z \leftarrow Gritty \rangle$ an x-variant of (71a)

So assuming the same domain, the sequence in (71a) fails to satisfy (70) because of the problematic x-variant in (71c). Intuitively, there's something in the domain that isn't green, so (70) isn't satisfied.

But it is true that all the frogs in the domain are green. So we could imagine a restricted version of (70) that uses the prefix "$\forall x(frog(x))$" instead of "$\forall x$".

Namely, (72) is satisfied by a sequence iff each x-variant of that sequence that assigns a frog to x also satisfies (70).

(72) $\forall x(frog(x))[green(x)]$

$\approx$ *everything that is a frog is green*

(73) a. $\langle x \leftarrow Kermit; y \leftarrow Mr.Toad; z \leftarrow Gritty \rangle$ satisfies (72)

  b. $\langle x \leftarrow Mr.Toad; y \leftarrow Mr.Toad; z \leftarrow Gritty \rangle$ an x-variant of (73a)

  c. $\langle x \leftarrow Gritty; y \leftarrow Mr.Toad; z \leftarrow Gritty \rangle$ ~~an x-variant of (73a)~~

This extra restriction – "$(frog(x))$" – means that the previously problematic x-variant in (73c) is not considered, because it does not assign a frog to x. After eliminating this x-variant, the sentence (72) is satisfied by the sequence (73a). In effect, the domain is restricted to the frogs (just as the universe of evaluation was restricted to the set of frogs in (67b)).

In a sense, the idea of restricted quantification goes back to Aristotle, who focused on syllogisms like (74) and (75) (see Parsons (2014) for review).

(74) No frog is orange

  Kermit is a frog

  $\therefore$ Kermit isn't orange

(75) Every frog is an animal

  Every animal is mortal

  $\therefore$ Every frog is mortal

For Aristotle, the propositions that serve as the premises and conclusions in the above syllogisms had a subject/predicate structure (see Chapter 2 of Pietroski (2018)). Predicates like *is orange*, *is a frog*, *is an animal*, and *is mortal* were treated as ways of categorizing different subjects. And quantificational terms like *every frog* and *no frog* played the same logical role as singular terms like *Kermit*. So determiners like *every* and *no* were not treated as tools for relating two predicates, but as terms that combine with a noun to yield a subject, to which the predicate then applies (see section 2.1.1 regarding quantifiers in object position).

In a translation of a commentary on Aristotle, Hodges (2012) puts the point like this: "Determiners ... combine with the subject terms and indicate how the predicate relates to the number of individuals under the subject; ... *Every man is an animal* signifies that *animal* holds of all individuals falling under *man*" (p.247).

With these three ways of thinking about restricted quantification in mind, we can capture the idea using the formalism introduced in Chapter 1. In particular, (76) is a possible representation for *every* that eschews relations in favor of restricted quantification.

(76) $\lambda \mathbf{I}.\lambda \mathbf{E}.\iota G(\forall x[Gx \equiv \mathbf{I}x])[\mathbf{E}(G)]$

$\approx$ *the things$_G$ that are $\boldsymbol{I}$ are such that they are $\boldsymbol{E}$*

Here, only the things that satisfy the $\mathbf{I}$ condition are grouped (under the variable G). The $\mathbf{E}$ predicate does not name an independent group, it supplies a further condition that the group G must meet. That is, the whole sentence obtains when the things that meet the $\mathbf{I}$ condition also meet the $\mathbf{E}$ condition. For example, *Every*

frog is green holds just in case the frogs all meet the additional condition of being green. The representation in (76) is thus not relational, but restricted.[2]

In sum, the main difference between relational and restricted quantification is in how the arguments are treated. For relational quantification, the quantifier's two arguments are logically on a par. The internal and the external arguments are both terms in a relation. The determiner specifies which relation. On the restricted view, the quantifier's arguments serve different logical roles. The internal argument restricts the domain of quantification in some sense, and the external argument supplies a further condition that some quantity of this restricted domain needs to meet. The determiner specifies the quantity.

Chapter 1 discussed some of the initially promising motivations for the relational view and how they are undercut by current developments in semantic theory. We saw in section 1.4.1 that the need for increased expressive power to deal with proportional quantifiers like *most* can be provided by restricted quantification as well as generalized quantifiers (and that the latter over-generates by design). And we saw in section 1.4.2 that event semantics tells against an analogy between verbs-as-relations and determiners-as-relations. Likewise, section 1.4.3 showed that restricted

---

[2] A representation could still be restricted in the relevant sense even if it relates two groups. For example, (77) is genuinely relational way to represent universal quantification, through the identity relation, but (78), which also uses identity, is restricted in the relevant sense.

(77) $\lambda\mathbf{I}.\lambda\mathbf{E}.\iota G(\forall x[Gx \equiv \mathbf{I}x])\{\iota G'(\forall y[G'y \equiv \mathbf{I}y \ \& \ \mathbf{E}y])[G = G']\}$ (relational)
     $\approx$ *the things$_G$ that are $\mathbf{I}$ are identical to the things$_{G'}$ that are $\mathbf{I}$ and $\mathbf{E}$*

(78) $\lambda\mathbf{I}.\lambda\mathbf{E}.\iota G(\forall x[Gx \equiv \mathbf{I}x])\{\iota G'(\forall y[G'y \equiv Gy \ \& \ \mathbf{E}y])[G = G']\}$ (restricted)
     $\approx$ *the things$_G$ that are $\mathbf{I}$ are identical to the things$_{G'}$ that are both one of them$_G$ and $\mathbf{E}$*

In particular, the second group (G') in (78) is not an independent group, but one that is in part defined by the first group, G. In other words, G' is restricted by G in (78) but is completely independent of G in (77).

quantification creates no mismatch between logical and grammatical form. After a brief discussion of dealing with quantifiers in object position (section 2.1.1), the remainder of this chapter offers some positive reasons for returning to the hypothesis that natural language determiners – or, at the least, that *every* and *all* – are mentally represented as devices for creating restricted quantifiers.

2.1.1 Quantifiers in object position

As discussed above, restricted quantification preserves the Aristotelian idea of a subject/predicate distinction. In a sentence like *every frog is green*, the subject *every frog* sets the domain and the predicate *is green* is predicated exhaustively of that restricted domain. Relational quantification jettisons this idea in favor of the Fregean alternative that both arguments – *frog* and *is green* – are logically on a par.

Of course, quantifiers like *every frog* can appear in sentences like *Kermit saw every frog*. This might at first seem like an issue for the restricted view: Can we retain the idea that *every frog* is the subject even when it appears inside of a predicate like *saw every frog*? In short, yes, assuming quantifier raising (Higginbotham and May 1981; May 1985).

In particular, assume that when the quantifier is in subject position, like in (79), *every*'s internal argument is *frog* and its external argument is t_i *is green*. Likewise, when the quantifier is in object position, like in (80), *every*'s internal argument is *frog* and its external argument is *Kermit saw* t_i. In both cases, t_i is the trace of displacement left by quantifier raising, which is co-indexed with the raised

quantifier *every frog.*

(79) Every frog is green

```
                        S
                       / \
                    DPᵢ    VP
                   / \     /\
                  D   NP   tᵢ is green
                  |    |
                Every  frog
```

(80) Kermit saw every frog
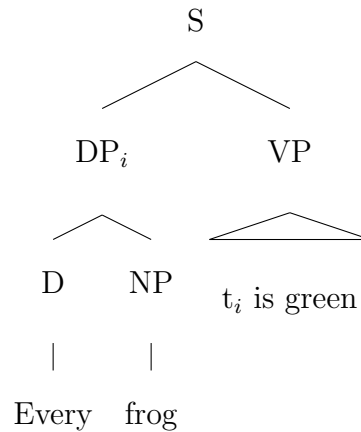
```
                        S
                       / \
                    DPᵢ    VP
                   / \     /\
                  D   NP   Kermit saw tᵢ
                  |    |
                Every  frog
```
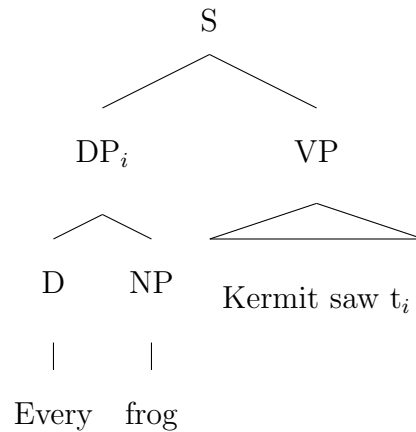
Given these logical forms, we can still treat *every frog* as a restricted quantifier – and consequently as the subject of the sentence – even when it initially occurs within a predicate. The indexed trace t_i is treated as a variable, so *Kermit saw t_i* is analogous to *Kermit saw it$_i$* or *Kermit saw x*. The trick is to say, as part of the rule of interpretation, that because t_i is co-indexed with the displaced quantifier, the values of this variable are restricted by *every*'s internal argument as well. That is, $t_i = x$.

Taking the approach of quantifying over sequences of assignments of values to

variables, (80) can be treated the same way as (72) from earlier. Namely, (80) is represented as (81).

(81) $\forall x(frog(x))[KermitSaw(x)]$

$\approx$ *everything that is a frog is s.t. Kermit saw it*

The whole sentence is TRUE, relative to an assignment A, just in case every x-variant of that assignment makes the embedded open sentence *Kermit saw x* TRUE. And the embedded sentence by itself is TRUE, relative to A, just in case Kermit saw the value assigned to x. Put another way, the values that x can take on are restricted to frogs, and the whole sentence is TRUE just in case *Kermit saw x* is TRUE for any value of x.

In sum, the fact that determiners often appear in object position creates no issues for the idea that they are devices for creating restricted quantifiers so long as quantifiers raise, as is standardly assumed (though see Jacobson 1999, 2014). We now turn to psycholinguistic and linguistic reasons for preferring the restricted view.

2.2 Psychosemantic evidence: Which arguments are represented?

Given the linking hypothesis discussed in section 1.3.1 of the previous chapter (Lidz et al. (2011)'s Interface Transparency Thesis), these two views make different predictions about how participants will verify quantificational sentences. If quantifiers like *every* have relational meanings (like (66), the GQT treatment), then participants should be biased to verify *every*-sentences by representing and relating both arguments. The relational representation in (82), for example, implicates two

independent groups, the big circles (G) and the blue things (G'), and relates them with "$\subseteq$".

(82) Every big circle is blue

$$= \iota G(\forall x[Gx \equiv \textbf{big-circle}(x)])\{\iota G'(\forall y[G'y \equiv \textbf{is-blue}(y)])[G \subseteq G']\}$$

This suggests a straightforward verification procedure: represent the big circles and the blue things and check whether the former is a subset of the latter. Behaviorally then, this predicts that participants will treat the two arguments symmetrically, as *big circles* and *is blue* play the same logical role in the semantic representation.

On the other hand, if quantifiers have restricted meanings that treat the arguments as logically asymmetric (like (76) from earlier), then participants should be biased to treat the extensions of those arguments as psychologically asymmetric. For example, the particular restricted representation in (83) implicates just one group, the big circles (G).

(83) Every big circle is blue

$$= \iota G(\forall x[Gx \equiv \textbf{big-circle}(x)])[\textbf{is-blue}(G)]$$

It does not call for treating the blue things as an independent group to which the big circles are related. This suggests a verification strategy that involves representing the big circles, then checking if they meet the condition supplied by the second predicate (i.e., checking if they are blue).

To be sure, there could also be a restricted representation that implicates more than one group – see note 2 above – but even so, the external argument, on its own, would not be predicted to be explicitly represented.

In terms of strategies they suggest then, the crucial difference between the relational (82) and the restricted (83) is whether the external argument is represented. The following seven experiments test this prediction. We will see that evaluating sentences with *every* or *all* causes participants to represent the determiner's internal argument, but not its external argument. This suggests that the semantic representations are restricted, not relational. We will also see, perhaps surprisingly, that participants do not represent the conjunction of the determiner's internal and external arguments (e.g., the big blue circles). So in addition to being restricted in the relevant sense, the semantic representations of *every* and *all* seem to implicate only a single group.

2.2.1   Cardinality knowledge as a proxy for argument representation

These experiments, some of which were first reported in Knowlton et al. (2021b), use cardinality knowledge as proxy for whether participants mentally represented a given argument during sentence evaluation. The idea is that if the argument is represented as an independent group, participants should be able to estimate the number of members in that group. This is consistent with how groups are understood in cognitive psychology, a topic discussed at length in Chapter 3.

In particular, participants in these experiments were shown dot-displays like the one in Figure 2.1. They were asked to evaluate sentences like *every big circle is blue* with respect to these displays. After answering TRUE or FALSE, the image disappeared, and participants were asked to recall the cardinality of some group of

circles from the display (e.g., *how many big circles were there?*).
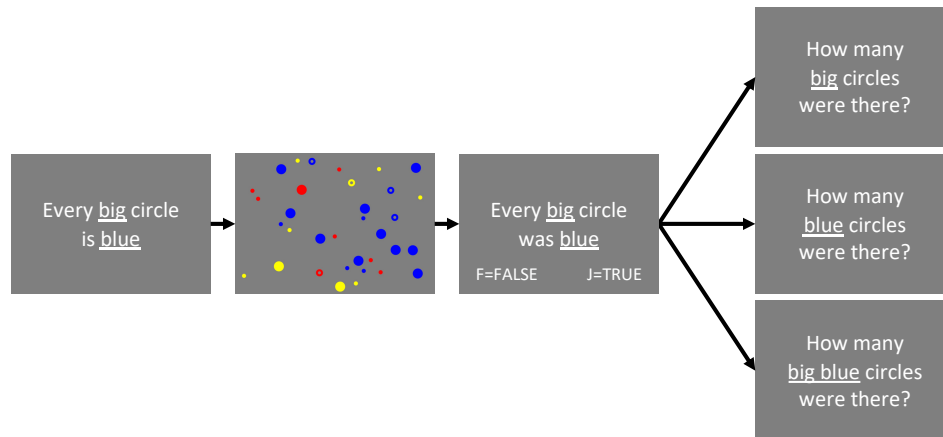
Figure 2.1: Trial structure of Experiment 1 (and subsequent experiments in this chapter). Some follow-up cardinality questions probed the internal argument (big), others probed the external argument (blue), and others probed the conjunction of the two (big blue).

Some follow-up cardinality questions targeted the internal argument (e.g., big circles, in this case), others targeted the external argument (e.g., blue circles), and others targeted the conjunction of both arguments (e.g., big blue circles). Participants' performance on these cardinality questions was compared to their performance on a baseline task, using the design in Halberda et al. (2006). Participants were shown the same sorts of dot-displays and asked to estimate the cardinality of some group from the display, but they were told which group to focus on prior to being shown the image. They might be asked, for example, *how many big circles are there?*, then view the image and submit their response. Performance on this "probe-before" baseline task represents the best possible cardinality estimates the visual system can afford, given the details of these displays. Moreover, comparing responses to a baseline task controls for the possibility that different groups are dif-

ferentially difficult (e.g., given the vibrant colors used, color is likely a more salient cue than size for grouping circles in these experiments).

To the extent that participants represent an argument, they should show similar cardinality-estimation performance on the baseline task and the sentence verification task. That is, if a participant is able to estimate the cardinality of, say, the big circles just as well after evaluating a sentence like *every big circle is blue* as they are when they are told ahead of time to estimate the cardinality of the big circles, it strongly suggests that they attended to and represented the big circles when evaluating the sentence. And since participants are able to enumerate up to three groups of circles in parallel (Halberda et al., 2006), there is no non-linguistic pressure that would prevent them from accurately estimating the cardinalities of the big circles and the blue circles if they had adopted a verification strategy that relied on representing and relating the internal and external arguments.

Experiment 1 (section 2.2.3) shows that after evaluating sentences with an internal argument defined by a size and an external argument defined by a color (e.g., *every big circle is blue*), participants know the cardinality of the group defined by that size (e.g., big) as well as their visual systems will allow. But after evaluating the same sentence, they do not know the cardinality of the group defined by that color (e.g., blue) or the cardinality of the group defined by the conjunction of both features (e.g., big blue). Experiment 2 (section 2.2.4) replicates this result in sentences with *all* instead of *every*. Experiment 3 (section 2.2.5) shows that the effect goes away when participants are asked to evaluate sentences in which the focus operator *only* replaces *every* or *all*, suggesting that the effect may be unique to quantificational

81

determiners.

Experiment 4 (section 2.2.6) replicates the finding that only the determiner's internal argument is represented using sentences in which the arguments are flipped such that the internal argument is defined by a color and the external argument is defined by a size (e.g., following *every blue circle is big*, participants know the cardinality of the blue circles but not the big circles or the big blue circles). In a similar vein, Experiment 5 (section 2.2.7) replicates the result with *only* using sentences that swap size and color in a similar way (e.g., evaluating *only blue circles are big* leads to below-baseline cardinality estimation performance across the board).

Experiment 6 (section 2.2.8) controls for the possibility that the way the predicates are introduced is responsible for the effect, by testing sentences in which the relevant portion of the internal argument is introduced by a relative clause, as in *every circle that is big is blue.*

Experiment 7 (section 2.2.9) then replicates the initial result in a paradigm that offers participants the ability to opt out of particular cardinality questions by pressing an "I don't know!" button. Participants opt out at higher-than-baseline rates when cardinality questions probe the external argument or the conjunction of the internal and external arguments, but they opt out at the baseline rate when asked about the internal argument.

2.2.2 Measuring cardinality knowledge

Accurately measuring cardinality knowledge requires understanding details of the cognitive system that underpins numerical estimation: the Approximate Number System (ANS; for review, see Feigenson et al. (2004) and Dehaene (2011)). The ANS is an evolutionarily ancient system for representing and comparing numerosities. The representations that this system generates are standardly modeled as Gaussian tuning curves ordered along an internal scale – the "mental number line" – as in Figure 2.2.[3] Perceived numerosities (e.g., a certain number of shapes presented on a computer monitor) "activate" different parts of the scale, which is linearly ordered such that larger numerosities are represented farther to the right. Numerical estimates of a given numerosity can be thought of as a sample from the relevant "activation pattern." So if the participant modeled by Figure 2.2 were shown nine items, they would most often respond "nine," less often respond "eight" or "ten," and less often still, "seven" or "eleven".

The standard deviation of these "activation patterns" increases linearly with the mean. The distribution resulting from being shown nine things, for example, is wider than the distribution resulting from being shown four things. This property of the model is known as scalar variability and it captures the fact that ANS representations are ratio-dependent (e.g., 9 and 10 are just as difficult to distinguish as 90 and 100, despite the latter having a larger absolute difference). It also explains why

---

[3] The ANS shares a representational format with many other systems. As such, this model can be applied to other psychological dimensions, like loudness, brightness, and distance (see Odic et al. (2016) for a helpful review).
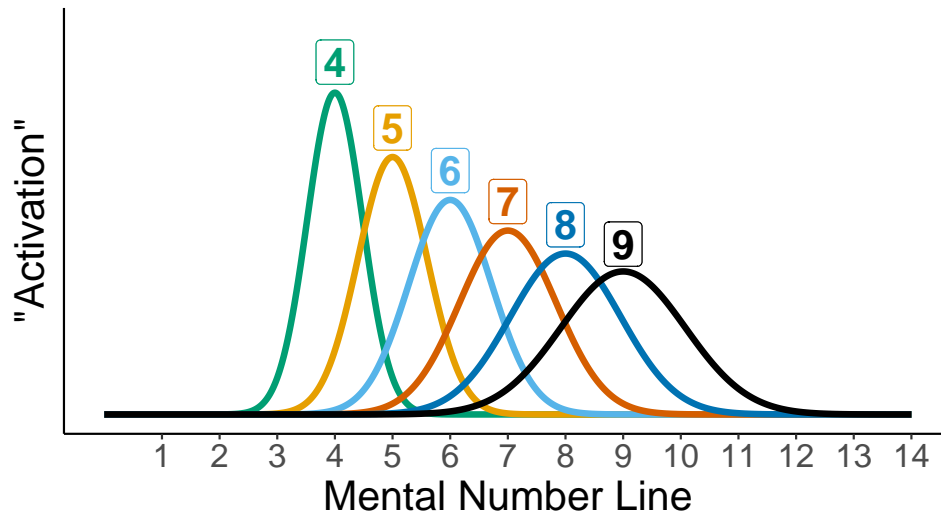
Figure 2.2: The standard model of Approximate Number System representations. Numerals in squares represent the numerosity presented (e.g., the number of shapes presented in an experiment), the distributions below each numeral represent the level of "activation" along each point of the mental number line.

participants are more confident about smaller numerosities: A narrower distribution means less overlap with nearby numbers on the mental number line.

The rate of increase of the standard deviation is subject to a good deal of individual variation, and individuals' precision correlates with their performance on the math portion of college-entrance exams (Libertus et al., 2012). It also develops throughout the lifespan, such that young adults have more precise ANS representations than children (Halberda and Feigenson, 2008) and peak in their precision around age 30 (Halberda et al., 2012). So, in practice, some individuals have narrower "activation patterns" than others (though both would, on average, offer the same response if shown some given numerosity many times).

Figure 2.2 is also an idealization in another important sense: The relevant

signal from the environment is often compressed in the process of building a representation of numerosity, leading to underestimation (Krueger, 1984; Odic et al., 2016; Stevens, 1964). That is, individuals usually show some degree of inaccuracy, even on average. The distribution of activation that results from being shown, say, nine things may be centered on a value below 9 on the mental number line, as in Figure 2.3.[4] So instead of answering "nine" on average, participants may systematically underestimate and answer "six". The degree of underestimation increases with the numerosity based on the psychophysical power law: $estimate = signal^\beta$, where β is the level of accuracy (in Figure 2.2, $\beta = 1$; in Figure 2.3, $\beta = 0.8$).

Taking both the internal "noise" and signal compression into account, numerical estimates (y) in response to some numerosity (x) are modeled as the Gaussian distribution in (84), where β is the "accuracy" parameter that represents the degree of signal compression, σ is the rate of growth of the standard deviation (i.e., an individual's precision), and α is a scaling factor.

$$(84) \quad y \sim \mathcal{N} \left(\begin{array}{l} mean = \alpha x^\beta \\ stdev = \sigma \times mean \end{array} \right)$$

In all of the experiments reported below, we use participants' accuracy – β – as the dependent measure. An accuracy closer to 1 reflects better cardinality estimation, whereas an accuracy below 1 reflects some degree of underestimation. As noted above, cardinality estimation accuracy on the critical sentence verification task can

---

[4] Instead of thinking of underestimation as reflecting signal compression, it may be thought of as reflecting a larger response bin for each numerosity (Izard and Dehaene, 2008). For example, the distribution of activation that results from being shown nine things may still be centered over 9 on the mental number line, but 9 on the mental number line may be incorrectly mapped to the verbal response "seven".
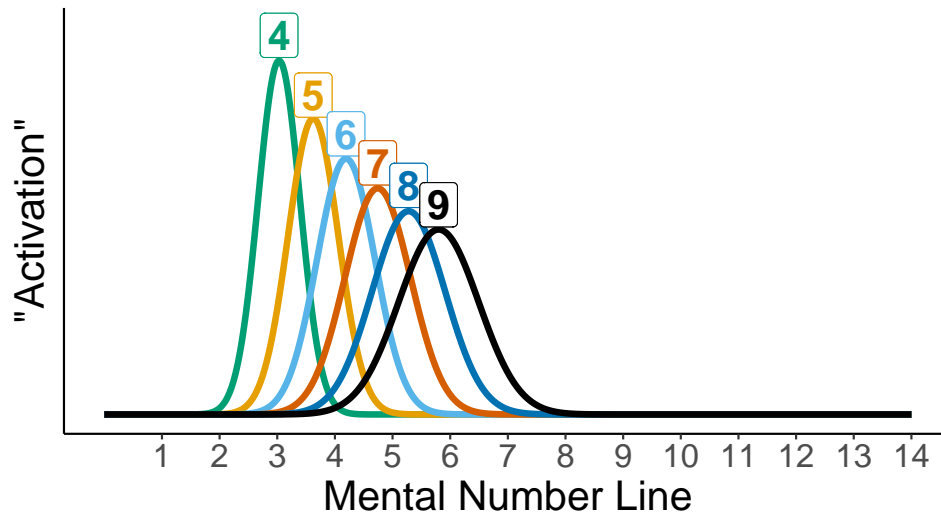
Figure 2.3: The standard model of ANS representations, given an internal β of .8. Numerals in squares represent the numerosity presented (e.g., the number of shapes presented in an experiment), the distributions below each numeral represent the level of "activation" along each point of the mental number line. Because accuracy is below 1, the signal is compressed (e.g., presenting nine items leads to a pattern of activation centered on a mental number line value below 9).

be compared against accuracy on a baseline task that indexes the visual system's best possible performance given these particular displays (i.e., the expected rate of signal compression).

2.2.3 Experiment 1: *every [size] circle is [color]*

Participants: 53 participants were recruited online using Amazon Mechanical Turk. All passed an English-screener prior to the actual task. Two participants were excluded from further analysis for performing at chance or below on the TRUE/FALSE portion of the experiment and three were excluded for taking longer than five seconds, on average, to respond to cardinality questions. This left 48

participants.

Materials: Sentences in the verification task had the form in (85).

(85) Every {big/medium/small} circle is {red/yellow/blue}

Half of these sentences were TRUE with respect to the display and half were FALSE.

Displays consisted of a grey background with red, yellow, and blue circles that could be big, medium, or small, as in Figure 2.1. Medium circles had grey holes in the middle, to make them more distinguishable from the other two sizes (Chen, 1982, 2005). The circle sizes and colors were labeled during the instructions to ensure that they could be correctly identified by participants.

Each display contained up to 48 circles of from the nine possible size/color combinations. Each size/color combination contained up to 10 circles, and there were always six size/color combinations for which at least one circle was present. On trials that were FALSE, there were between one and three disconfirming circles and most of the circles confirmed the statement (i.e., there were no cases where the sentence to evaluate was *every big circle is blue* and the display contained no big blue circles). The baseline cardinality estimation task used images generated with these same constraints.

Follow-up cardinality questions in the verification task were distributed in the following way. Four probed the target size (e.g., big circles if the initial sentence was *every big circle is blue*). Four probed the target color (e.g., blue circles). Four probed the target size/color combination (e.g., big blue circles). Six filler questions were included as well: Two probed a distractor size (e.g., small circles), two probed

a distractor color (e.g., yellow circles), and two probed a distractor conjunction (e.g., medium red circles).

Procedure: Participants first completed 15 trials of the baseline cardinality estimation task. They were first presented with a question that named a color, size, or color/size combination (e.g., *how many big circles are there?*). After pressing "space," the circles were displayed for one second, followed by a reiteration of the question (e.g., *how many big circles were there?*). Participants could then type their answer and move on to the next trial.

After completing the baseline task, participants were then given 18 trials of the verification task. They were presented with a statement (e.g., *every big circle is blue*) and after pressing "space," were shown a display of circles for one second. The statement then returned to the screen and participants evaluated it as TRUE or FALSE relative to the display by pressing "J" or "F" on their keyboard. After offering their TRUE/FALSE response, they were given a follow-up cardinality question (e.g., *how many big circles were there?*) and typed in their guess before moving to the next trial.

Predictions: As discussed above, relational and restricted meanings suggest different verification strategies. If *every* has a relational meaning, then both arguments should be treated on a par, so when evaluating a sentence like *every big circle is blue*, participants should be biased to represent and relate the big circles and the blue circles.[5] As a consequence, participants should have high accuracy when asked

---

[5] More accurately, the relational view predicts that they should represent the blue things. But in this experimental context, all of the relevant things are circles.

about the cardinality of a group defined by the target size or the target color. That is, they should be close to their baseline accuracy for the relevant type of group.

In contrast, if *every* has a restricted meaning, the arguments should be treated asymmetrically. In particular, the things picked out by the external argument (the blue circles) should not be explicitly represented. As a consequence, participants should have lower accuracy on the verification task than baseline when asked about the cardinality of a group defined by the target color. On the other hand, the things picked out by the internal argument (the big circles) should be explicitly represented if the representation proposed in (83) is the meaning of *every big circle is blue*.[6] Participants should thus have high accuracy (relative to baseline) when asked about the cardinality of the group of circles picked out by the target size.

Predictions are less clear when it comes to the conjunction of both target features (e.g., the big blue circles). A representation might still be restricted even if it implicates a group whose members are the satisfiers of both the predicate supplied by the internal argument and the predicate supplied by the external argument (see note 2). Of course, there are could also be a genuinely relational representation that relates the big circles and, as an independent group, the big blue circles in a sentence like *every big circle is blue* (see section 2.3.3 for a proposal along these lines). So

---

[6] The prediction of restricted quantification in general is that the internal argument will be represented to the exclusion of the external argument, not that the internal argument will be mentally grouped. It is also possible for a representation to be restricted but to not predict that the things picked out by the internal argument will be represented in a way that supports cardinality estimation. In particular, the representation in question might lack any variables that group the satisfiers of a predicate together (e.g., the first-order "$\lambda \mathbf{I}.\lambda \mathbf{E}.\forall x(\mathbf{I}x)[\mathbf{E}x]$"). This possibility is discussed at length in Chapter 3, which argues that while a restricted second-order representation is the meaning of *every*, a restricted first-order representation is the meaning of *each*. This makes the prediction that *each*-sentences, unlike *every*-sentences, will lead participants to represent the internal argument as individual objects, but like *every*-sentences, that the external argument will not be explicitly represented.

finding that participants accurately recall the cardinality of the conjunction group would leave the relational or restricted question unsettled.

On the other hand, finding that participants are able to recall the cardinality of the group named by the internal argument but unable to recall the cardinality of the conjunction group would constitute a striking result. After all, whenever the sentence under evaluation is TRUE, the answers to these two cardinality questions will be identical. If there are nine big circles and it's TRUE that every big circle is blue, then there are nine big blue circles. But if participants represent these *every*-sentences as the restricted second-order single-group representation in (83), then the conjunction group will not be explicitly implicated in their mental representation of the sentence. Consequently, they should not be expected to have a good estimate of its cardinality readily available in memory (though they may be able to reason their way to the answer if given enough time to reflect on it). Such a result would offer strong evidence in favor of the particular restricted representation in (83).

Results: On the TRUE/FALSE portion of the verification task, participants correctly evaluated 74% of the sentences. On the cardinality estimation portion, their responses, plotted in Figure 2.4, bear out the predictions of the restricted quantification hypothesis.

In particular, a difference score for each question type was calculated by taking participants' cardinality estimation accuracy following sentence verification and subtracting their baseline cardinality estimation accuracy. Each difference score was compared against 0 with a Wald test, so a significant difference indicates that participants were worse than their visual system would allow when asked to give

90

Figure 2.4: Cardinality estimation difference scores from Experiment 1. A difference score of 0 reflects best possible performance given these displays. Accuracies when asked about groups named by size, color, or both after evaluating sentences of the form *every [size] circle is [color]* were calculated and subtracted from accuracy in a baseline number estimation task using the same dot-displays. Error bars reflect the square root of the sum of squared errors around both accuracies.

cardinality estimates of the relevant circles. We find that participants know the cardinality of the group picked out by the internal argument (e.g., big circles): their difference score is not significantly lower than 0 ($\chi^2 = 0.007, p = .93$). However, participants were significantly less accurate than baseline when asked about the cardinality of the group picked out by the external argument (e.g., blue circles) ($\chi^2 = 13.13, p < .001$) and when asked about the group picked out by the relevant conjunction of features (e.g., big blue circles) ($\chi^2 = 27.09, p < .001$).

As noted above, the conjunction result is especially impressive. It shows that after evaluating a sentence like *every big circle is blue* participants know the cardinality of the big circles but not the cardinality of the big blue circles. In half the

trials of the experiment, the sentence under evaluation was TRUE so the cardinality of these two groups was identical. Even so, considering just the TRUE trials, we find the same result, as seen in Figure 2.5. Participants knew the cardinality of the group picked out by the internal argument ($\chi^2 = 0.205, p = .65$) but not the conjunction group ($\chi^2 = 15.41, p > .001$) (and, not the cardinality of the group picked out by the external argument ($\chi^2 = 9.87, p < .01$)).
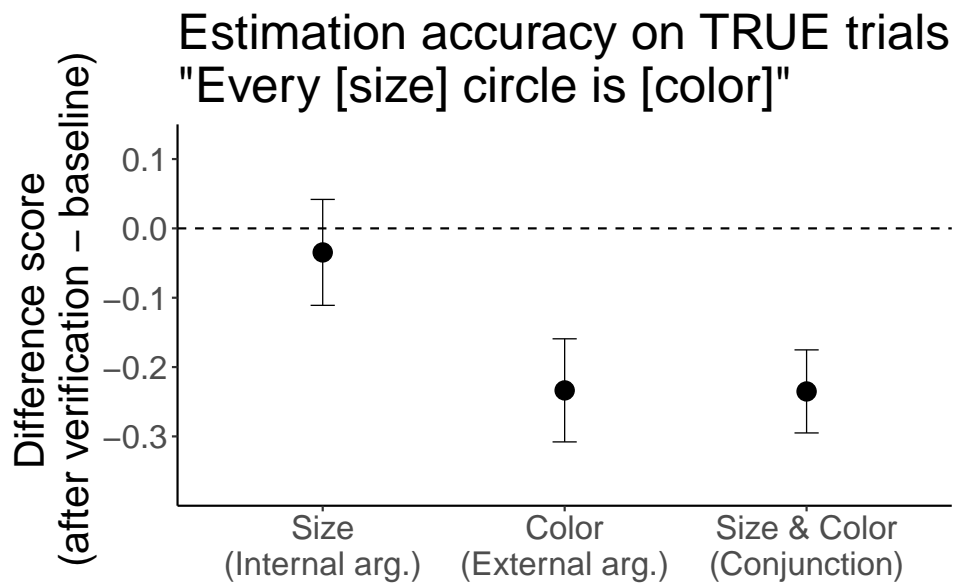
Figure 2.5: Cardinality estimation difference scores from Experiment 1 on the TRUE trials. In these cases, the correct answer when asked about the internal argument and the conjunction of the internal and external arguments was identical. A difference score of 0 reflects best possible performance given these displays. Accuracies when asked about groups named by size, color, or both after evaluating sentences of the form *every [size] circle is [color]* were calculated and subtracted from accuracy in a baseline number estimation task using the same dot-displays. Error bars reflect the square root of the sum of squared errors around both accuracies.

In sum, after evaluating a sentence like *every big circle is blue*, participants know the cardinality of the big circles as well as their visual systems will allow; just as well as when they were instructed ahead of time to attend to and enumerate

the big circles. But after evaluating the same sentence and seeing the same image, participants were unable to accurately report the cardinality of the blue circles or the cardinality of the big blue circles. These results suggest that participants are only representing the internal argument, in line with the predictions of *every* having a restricted meaning, not a relational one. Moreover, the lack of representation of the conjunction of internal and external arguments provides support for the particular restricted representation that only implicates a single group. And as noted in section 2.2.1, this pattern of performance is likely not due to a constraint from the visual system: Halberda et al. (2006) show, using a similar task, that participants are able to enumerate up to three groups of circles in parallel with no additional working memory cost.[7]

While these results support the proposed restricted meaning for *every*, it might be that *every* is an outlier in this respect and that other quantifiers are relational. As a step toward answering the question of how far to generalize the restricted quantification hypothesis, Experiment 2 substitutes *every* with *all*.

2.2.4 Experiment 2: *all [size] circles are [color]*

Participants: 50 participants were recruited online using Amazon Mechanical Turk. All passed an English-screener prior to the actual task. One participant was excluded from further analysis for performing at chance on the TRUE/FALSE portion

---

[7] One thing that has not yet been empirically tested is whether participants are able to enumerate, in parallel, multiple groups with overlapping membership. In Halberda et al. (2006)'s displays, each group was defined by a different color, meaning each group was composed of distinct members. Here, membership overlaps: A small red circle counts toward the number of small circles, red circles, and small red circles.

of the experiment and one was excluded for taking longer than five seconds, on average, to respond to cardinality questions. This left 48 participants.

Materials: Materials were identical to Experiment 1 except that *every* was replaced by *all*, so sentences had the form in (86).

(86) All {big/medium/small} circles are {red/yellow/blue}

Procedure: The procedure was identical to that of Experiment 1.

Predictions: If *all* is like *every* in having a restricted meaning, we should observe the same results: high accuracy when asked to estimate the cardinality of a group defined by size and low accuracy when asked to estimate the cardinality of a group defined by color.

Results: As seen in Figure 2.6, this prediction was borne out: The same pattern of results is observed in Experiment 2 as was observed in Experiment 1. In particular, participants' difference score did not significantly differ from 0 on cardinality questions that probed the target size ($\chi^2 = 0.788, p = .375$). But participants were significantly less accurate than baseline when a cardinality question probed the target color ($\chi^2 = 24.67, p < .001$) or the relevant combination of size and color ($\chi^2 = 25.44, p < .001$). Less importantly, their performance on the TRUE/FALSE portion of the verification task was comparable as well (74.1% correct).

This replication with a second quantifier suggests that a restricted representation is not unique to *every*. Whether it generalizes beyond universal quantifiers to other determiners will be an interesting avenue for future research. As discussed in section 2.3, all determiners having restricted representations as their meanings
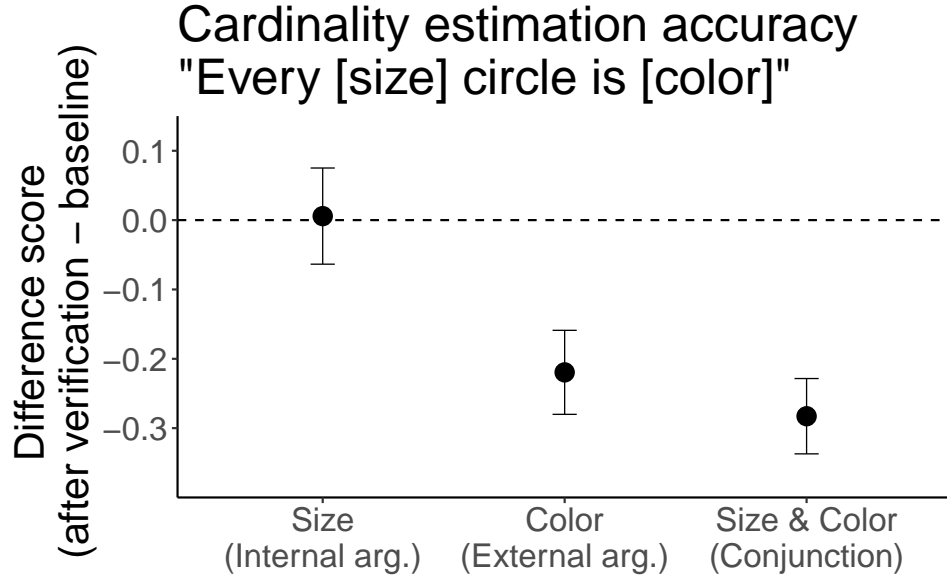
Figure 2.6: Cardinality estimation difference scores from Experiment 2. A difference score of 0 reflects best possible performance given these displays. Accuracies when asked about groups named by size, color, or both after evaluating sentences of the form *all [size] circles are [color]* were calculated and subtracted from accuracy in a baseline number estimation task using the same dot-displays. Error bars reflect the square root of the sum of squared errors around both accuracies.

would have important linguistic consequences.

In the meantime, alternative explanations for the present results loom large. Perhaps, for example, participants always represent the first noun phrase they encounter in a sentence, regardless of the sentence's semantic representation. To rule out this possibility, Experiment 3 replaces the determiner with the focus operator *only*, which does not take arguments like determiners.

2.2.5   Experiment 3: *only [size] circles are [color]*

Participants: 52 participants were recruited online using Amazon Mechanical Turk. All passed an English-screener prior to the actual task. Four participants were

excluded from further analysis for performing at chance or below on the TRUE/FALSE portion of the experiment and 1 was excluded for taking longer than five seconds, on average, to respond to cardinality questions. This left 47 participants.

Materials: Materials were identical to Experiment 1 except that *every* was replaced by *only*, so sentences had the form in (87).

(87) Only {big/medium/small} circles are {red/yellow/blue}

Procedure: The procedure was identical to that of Experiments 1 and 2.

Predictions: If the results of Experiments 1 and 2 arise because *[size] circle(s)* is the first argument of a determiner that has a restricted meaning, then those results should not replicate here. This is because *only* is not a quantificational determiner, but a focus operator (see Herburger (2000) for discussion and comparison with *even*). Unlike a determiner, *only* can be inserted at any point in a sentence like (88) (cf. *every*, which could only be added by removing a *the*).

(88) The cat thought that the dog found the bone

Moreover, *only* is focus-sensitive in a way that determiners like *every* are not. For example, (89a), with focus on *coffee*, cannot describe a situation in which students also ordered tea or soda. Likewise, (89b), with focus on *ordered*, cannot describe a situation in which students also brewed or paid for coffee.

(89)   a. Students only ordered coffee$_F$

      b. Students only ordered$_F$ coffee

(90)   a. Every student ordered coffee$_F$

b. Every student ordered$_F$ coffee

But the *every*-variants in (90) say nothing about the alternatives: (90a) may describe a situation where every student ordered coffee and some also ordered tea. And (90b) may describe a situation where every student ordered and paid for their coffee. So even though the experimental sentences in (87) look similar to the *every*- and *all*-sentences used in Experiments 1 and 2, they are not sentences in which *[size] circles* serves as the internal argument of a determiner.

This allows *only*-sentences to serve as a control. If the results of Experiments 1 and 2 merely arise because participants always represent the first NP they encounter in a way that supports cardinality estimation, and have nothing to do with the semantic representation of the sentence, then parallel results should be observed here. That is, participants should have high accuracy when asked to enumerate a group defined by a size, but low accuracy when asked to enumerate a group defined by a color or the conjunction of both features.

Results: Contra the predictions of this deflationary account, participants showed low cardinality estimation accuracy across the board, as seen in Figure 2.7. They performed significantly worse than baseline on questions probing size ($\chi^2 = 10.67, p < .01$), color ($\chi^2 = 62.08, p < .001$), or a size/color combination ($\chi^2 = 14.99, p < .001$).

This poor performance on the cardinality estimation portion of the task does not reflect overall task difficulty, as participants showed a similar level of accuracy at evaluating the initial TRUE/FALSE question as in Experiments 1 and 2 (73.6%).
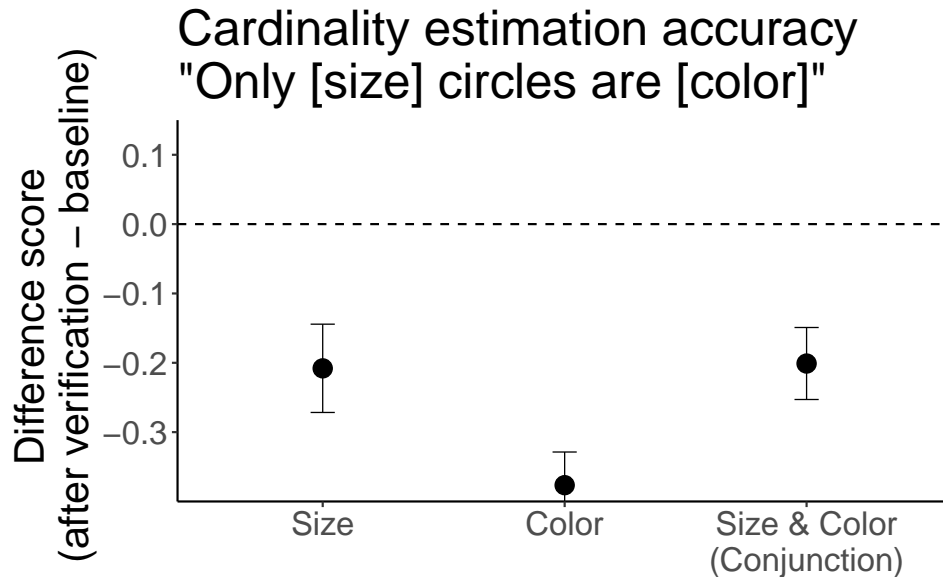
Figure 2.7: Cardinality estimation difference scores from Experiment 3. A difference score of 0 reflects best possible performance given these displays. Accuracies when asked about groups named by size, color, or both after evaluating sentences of the form *only [size] circles are [color]* were calculated and subtracted from accuracy in a baseline number estimation task using the same dot-displays. Error bars reflect the square root of the sum of squared errors around both accuracies.

These results militate against an alternative explanation for the results of Experiments 1 and 2: participants simply represent the first NP they encounter in these sorts of experimental contexts. Instead, it seems that there is something important about being the internal argument of a determiner. Of course, in both Experiments 1 and 2, this argument was defined by size. Experiment 4 further generalizes the paradigm by switching the internal and external arguments.

2.2.6 Experiment 4: *every [color] circle is [size]*

Participants: 51 participants were recruited online using Amazon Mechanical Turk. All passed an English-screener prior to the actual task. Three participants

were excluded for taking longer than five seconds, on average, to respond to cardinality questions. This left 48 participants.

Materials: Materials were identical to Experiment 1 except that the internal argument was defined by color and the external argument was defined by size. This meant that sentences had the form in (91).

(91) Every {red/yellow/blue} circle is {big/medium/small}

Procedure: The procedure was identical to that of Experiments 1-3.

Predictions: As noted above, in both Experiments 1 and 2 the internal argument was defined by a size and participants had accurate cardinality estimates of groups defined by that size. But to the extent that the results reflect restricted quantification, they should flip when the arguments are flipped: If the internal argument is defined by a color and the external argument is defined by a size, participants should instead represent the group defined by the relevant color.

Results: Consistent with this prediction, participants did not significantly differ from baseline when asked a cardinality question probing target color ($\chi^2 = 2.53, p = .112$), but they were significantly less accurate when asked about the target size ($\chi^2 = 16.16, p < .001$) or the relevant conjunction of color and size ($\chi^2 = 23.69, p < .001$). These results can be seen in Figure 2.8.

Participants seemed to find this task easier than the task in Experiments 1 and 2, as evidenced by their higher accuracy on the TRUE/FALSE portion of the verification task (86.3% correct). This is likely because color is a more salient cue than size in these stimuli (see Experiment 7 in section 2.2.9 for corroboration of this
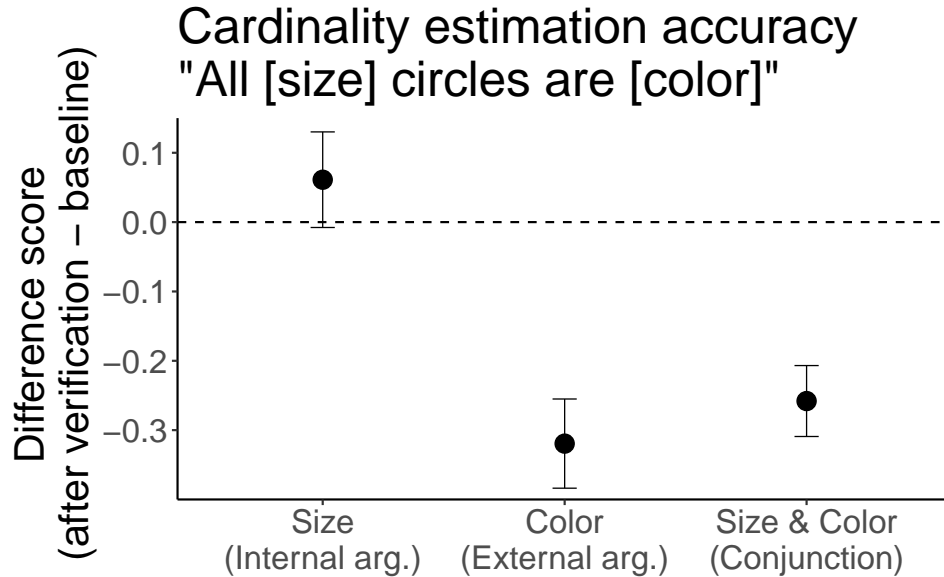
99
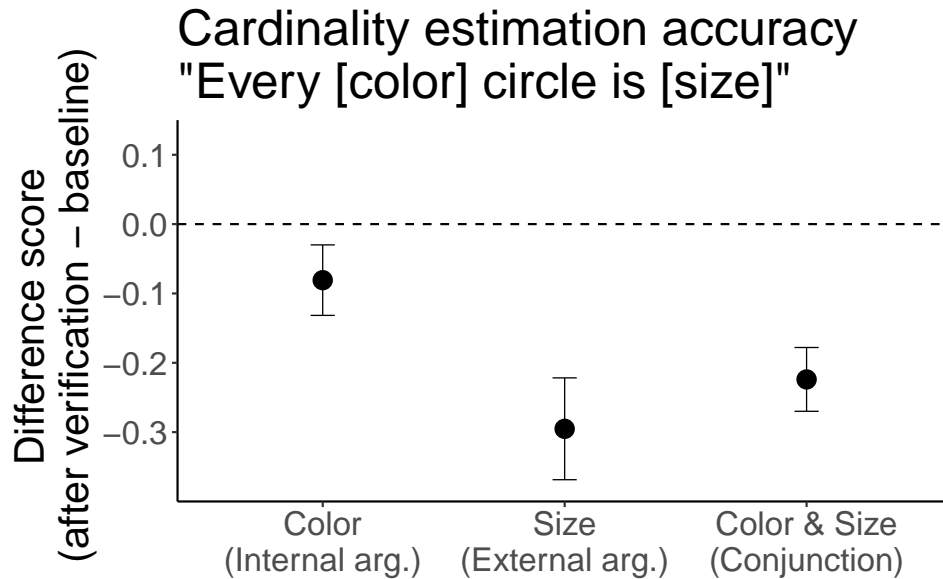
Figure 2.8: Cardinality estimation difference scores from Experiment 4. A difference score of 0 reflects best possible performance given these displays. Accuracies when asked about groups named by size, color, or both after evaluating sentences of the form *every [color] circle is [size]* were calculated and subtracted from accuracy in a baseline number estimation task using the same dot-displays. Error bars reflect the square root of the sum of squared errors around both accuracies.

point).

In any case, the present results further confirm that the results in Experiments 1 and 2 did not reflect something special about size. That is, given the poor accuracy for size questions in Experiments 3 and 4, it is not the case that participants always represent groups defined by size given these displays. Instead, they seem to represent the internal argument of determiners. To complete the paradigm, Experiment 5 again tests *only*-sentences, but with the order of the NPs flipped.

2.2.7 Experiment 5: *only [color] circles are [size]*

Participants: 55 participants were recruited online using Amazon Mechanical Turk. All passed an English-screener prior to the actual task. Six participants were excluded from further analysis for performing at chance or below on the TRUE/FALSE portion of the experiment and one was excluded for taking longer than five seconds, on average, to respond to cardinality questions. This left 48 participants.

Materials: Materials were identical to Experiment 3 except that the color was mentioned first and size was mentioned second (as in Experiment 4). This meant that sentences had the form in (92)

(92) Only {red/yellow/blue} circles are {big/medium/small}

Procedure: The procedure was identical to that of Experiments 1-4.

Predictions: Given the conclusions drawn from Experiments 3 and 4, participants here should show a lack of cardinality knowledge across the board.

Results: In line with this prediction, participants were significantly less accurate than baseline when asked about a color ($\chi^2 = 30.15, p < .001$), size ($\chi^2 = 5.14, p < .05$), or color/size combination ($\chi^2 = 8.34, p < .01$), as seen in Figure 2.9. Accuracy on the TRUE/FALSE portion was comparable to that in the preceding experiments (73%).

This experiment serves as a conceptual replication of Experiment 3. While evaluating sentences with *every* or *all* drives participants to mentally represent the internal argument, evaluating similar (at least on the surface) sentences with *only*
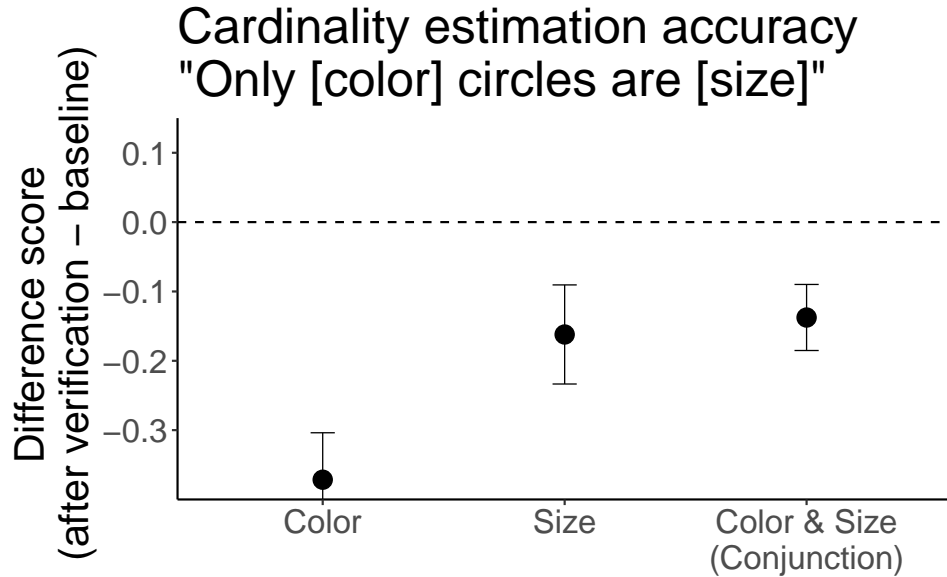
Figure 2.9: Cardinality estimation difference scores from Experiment 5. A difference score of 0 reflects best possible performance given these displays. Accuracies when asked about groups named by size, color, or both after evaluating sentences of the form *only [color] circles are [size]* were calculated and subtracted from accuracy in a baseline number estimation task using the same dot-displays. Error bars reflect the square root of the sum of squared errors around both accuracies.

does not lead to the same behavior.

Still, a remaining concern is that in the quantificational sentences used in Experiments 1, 2, and 4, there is a surface-level asymmetry between the two predicates. The internal argument is introduced as an NP (e.g., *big circle* or *blue circle*) whereas the external argument is introduced as a VP (e.g., *is big* or *is blue*). Perhaps the differences in how these predicates are introduced plays a role in driving attention to the internal but not the external argument. To control for this possibility, Experiment 6 equates the predicates by introducing the relevant part of the internal argument with a relative clause.

2.2.8 Experiment 6: *every circle that is [size] is [color]*

Participants: 53 participants were recruited online using Amazon Mechanical Turk. All passed an English-screener prior to the actual task. Five participants were excluded from further analysis for performing at chance or below on the TRUE/FALSE portion of the experiment. This left 48 participants.

Materials: Materials were identical to Experiment 1 except that the relevant predicate in the first argument was introduced with a relative clause, so sentences had the form in (93).

(93) Every circle that is {big/medium/small} is {red/yellow/blue}

In these sentences, the internal argument is *circle that is {big/medium/small}*. By adding a relative clause, the relevant predicate – *is {big/medium/small}* – is introduced in the same way, on the surface, as the relevant predicate in the external argument (*is {red/yellow/blue}*). In this way, the two arguments are equated in terms of their surface syntax.[8]

Procedure: The procedure was identical to that of Experiments 1-5.

Predictions: If the surface-level differences are responsible for the asymmetry in terms of how participants treat the two arguments, then the result should disappear as both predicates are introduced in the same way (e.g., *is big* for the internal argument and *is blue* for the external). But if a restricted meaning is responsible for the asymmetry, then this manipulation fail to make a difference. We should find

---

[8] Another way to achieve this would be to test sentences of the form *Every big one is a blue one*, given contextual support that *one* is anaphoric to circle. This version is currently underway.

the same effect as in Experiment 1: Participants should have high accuracy on size questions and low accuracy on color and conjunction questions.

Results: Results from the follow-up cardinality questions are plotted in Figure 2.10. As predicted, there was no significant difference between baseline performance and post-verification performance on questions that probed size ($\chi^2 = 0.407, p = .524$). Moreover, participants did not know the cardinality of the group picked out by the external argument, color ($\chi^2 = 23.41, p < .001$), nor did they know the cardinality of the group picked out by the conjunction of both target features ($\chi^2 = 11.16, p < .001$). On the TRUE/FALSE portion of the verification task, participants correctly evaluated 74.8% of the sentences, similarly to those in Experiment 1.
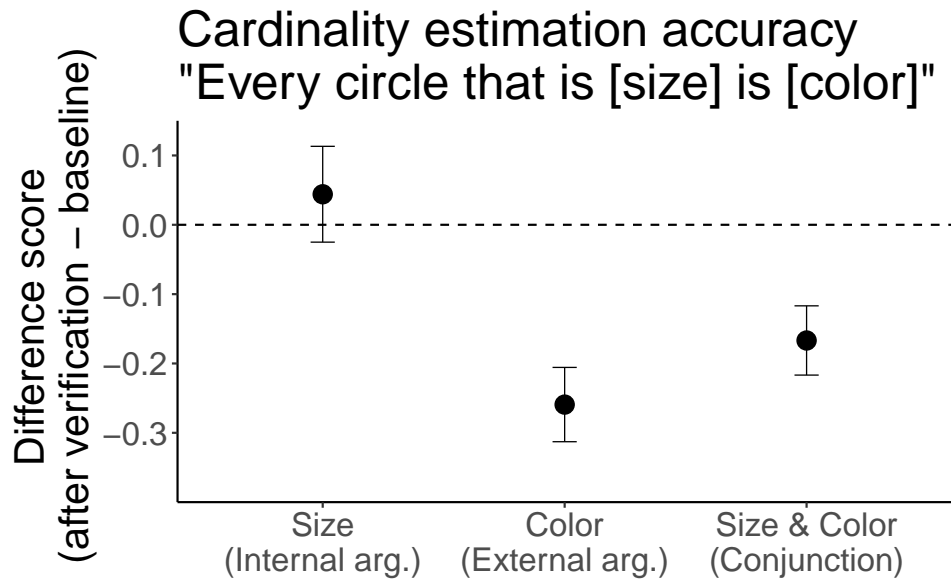
Figure 2.10: Cardinality estimation difference scores from Experiment 6. A difference score of 0 reflects best possible performance given these displays. Accuracies when asked about groups named by size, color, or both after evaluating sentences of the form *every circle that is [size] is [color]* were calculated and subtracted from accuracy in a baseline number estimation task using the same dot-displays. Error bars reflect the square root of the sum of squared errors around both accuracies.

Aside from providing a conceptual replication of Experiment 1, the relative clause manipulation controls for surface-level differences in how the two predicates are introduced. Despite this, we still see the same effect: participants know the cardinality of the group picked out by the determiner's internal argument, but not the the cardinality of the group picked out by the determiner's external argument.

As always, there are still possible alternative explanations that await further empirical exploration. To take one example, the internal argument in the above experiments was always in subject position. Knowlton et al. (2021b) report a version where the external argument is in subject position instead, though the results were inconclusive: While participants were well below baseline for cardinality questions targeting the external argument, they were also below baseline for questions targeting the internal argument. Future work will further explore the role of syntactic position (including information structural concerns like topichood). In the meantime, Experiment 7 offers another conceptual replication using a different dependent measure: rate of opting out of cardinality questions.

2.2.9   Experiment 7: Adding an "I don't know!" button

Participants: 56 participants were recruited online using Amazon Mechanical Turk. All passed an English-screener prior to the actual task. Four participants were excluded from further analysis for performing at chance or below on the TRUE/FALSE portion of the experiment and four were excluded for taking longer than five seconds, on average, to respond to cardinality questions. This left 48 participants.

Materials: The materials were identical to Experiment 1, with one exception: On both the baseline task and the sentence verification task, participants were given the option to opt out of answering the relevant cardinality question. Underneath the question (e.g., *How many big circles were there?*) there was a large red button labeled "I don't know!" Participants could, without penalty, press this button instead of making a guess.

Procedure: The procedure was identical to that of Experiments 1-6.

Predictions: In this version of the task, the main dependent measure is the rate of opting out (i.e., pressing the "I don't know!" button). The rate of opting out during the baseline task serves as an indication of how visually difficult it is to enumerate each type of group. Following sentence verification, if participants represent a particular group, they should opt out no more often than baseline. If they do not represent that group, they should opt out more often than the baseline rate. Building off of the past results then, participants here should be more likely than baseline to opt out given a question that probes a group picked out by the external argument (color) or by the conjunction of both arguments (a size/color combination).

Results: Participants' rate of pressing the opt-out button is given in Figure 2.11. The relatively higher baseline rate of opting out for size compared to color suggests that color is a more salient cue in these stimuli. The low baseline rate of opting out of answering conjunction questions may be due to the fact that, on average, these questions probed smaller cardinalities.

Importantly though, participants were no more likely than baseline to opt out

106
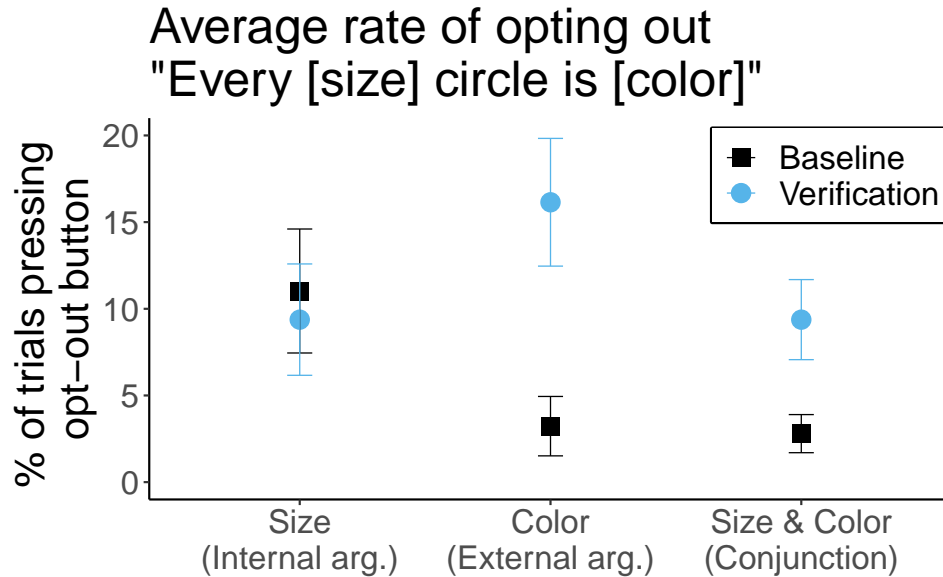
Figure 2.11: Participants' average rate of pressing the opt-out button labeled "I don't know!" instead of offering a cardinality estimate during the baseline task (black squares) and following sentence verification (blue circles).

of size questions ($t_{47} = 0.47, p = .64$). On the other hand, they were significantly more likely than baseline to opt out of questions probing color ($t_{47} = 3.07, p < .01$) or a size/color combination ($t_{47} = 2.91, p < .01$). This is a confirmation of the original result: Participants know the cardinality of the group defined by the internal argument as well as their visual system allows, but don't know the cardinalities of the groups defined by the external argument or the conjunction of both arguments. As a result, they opt out of answering more often than their baseline rate for those latter two types of questions.

Fitting the models to only the trials in which participants did not opt out yields the results in Figure 2.12. They were not significantly less accurate than baseline on size questions ($\chi^2 = 0.94, p = .332$) but they were significantly less accurate on color questions ($\chi^2 = 19.83, p < .001$) and conjunction questions ($\chi^2 = 4.19, p < .05$).

So even excluding the trials where participants knew they did not have a good estimate of the cardinality (and consequently opted out of answering), we find that participants still have poor estimates of cardinality when the external argument or the conjunction of both arguments is probed.
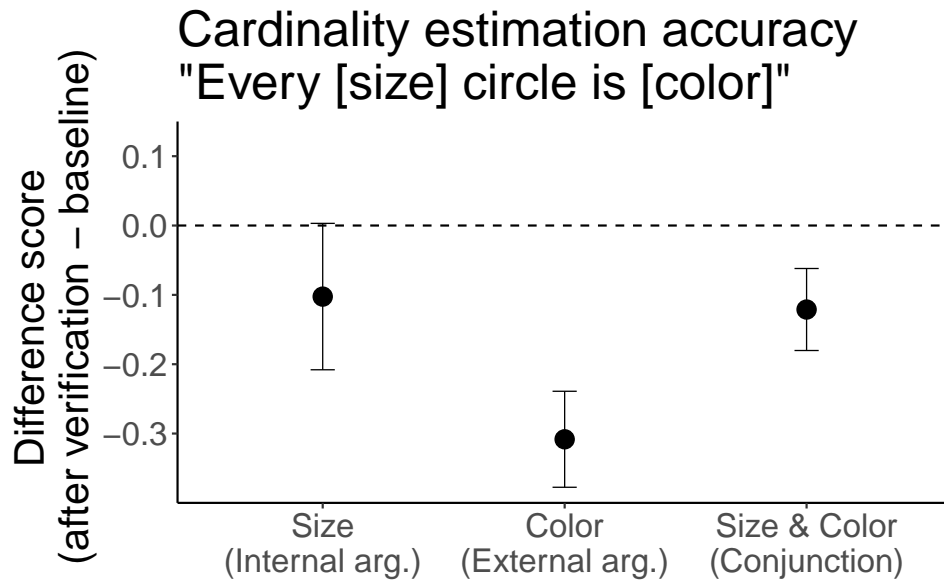
Figure 2.12: Cardinality estimation difference scores from Experiment 7. A difference score of 0 reflects best possible performance given these displays. Accuracies when asked about groups named by size, color, or both after evaluating sentences of the form *every [size] circle is [color]* were calculated and subtracted from accuracy in a baseline number estimation task using the same dot-displays. Error bars reflect the square root of the sum of squared errors around both accuracies.

These results reveal a new, straightforward signature of the effect. Interestingly, they also suggest that participants are only somewhat aware of their epistemic limitations. In any case, rates of opting out of cardinality questions and performance on cardinality questions not opted out of both point in the same direction: Participants only know the cardinality of groups picked out by the determiner's internal argument.

2.2.10 General discussion

Taken together, the results of these seven experiments suggest that participants explicitly represent *every*'s internal argument but not its external argument (and likewise for *all*). This would be surprising if *every* had a genuinely relational meaning. Why would participants avoid representing both arguments? As discussed, there is no reason to suspect a constraint from the visual system, so the apparent fact that participants treat the two arguments asymmetrically calls for explanation. On the other hand, these results are well-explained if *every* (along with *all*) has a restricted meaning that only calls for grouping the quantifier's internal argument.

Moreover, the results suggest that participants do not explicitly represent the conjunction of *every*'s (or *all*'s) internal and external arguments. This supports the particular restricted representation in (83), which implicates only a single group. It also provides an example where the semantic representation can have a powerful and unexpected effect on thought. Participants do not encode and remember every element of a scene; they focus their attention to the the particular group(s) highlighted by the meaning. So even in cases where two group descriptions are extensionally equivalent (e.g., the big circles and the big blue circles in a case where *every big circle is blue* is TRUE), participants may represent the relevant objects under one description (big circles) but not under the other (big blue circles).

As always, possible alternative explanations remain. Information structural concerns are of particular interest for future research to explore, as noted above. It might be, for example, that representation of a given argument is dependent on it

being the sentence topic.

It also remains a logical possibility that participants understand *every* to express a genuine relation between two groups, but nonetheless adopt a strategy, in these experiments, of only mentally grouping the extension of the internal predicate. But given that the visual system can represent three groups in parallel with no apparent cost to performance (Halberda et al., 2006), there is no obvious sense in which representing only one is an easier strategy. There is also no sense in which it is a superior strategy for this task. Participants know that a follow-up cardinality question is on the horizon, and Experiment 7 shows that they are at least somewhat conscious of their epistemic limitations. A two-group comparison strategy would lead to better performance on those cardinality questions, so it would be sensible for participants to adopt one.[9] Instead, they seem to avoid representing two groups. This would be hard to explain if *every* had a relational meaning.

2.3   Semantic evidence: Explaining "conservativity"

In addition to the psychosemantic evidence presented above, there is at least one compelling semantic reason to prefer restricted over relational quantification. As mentioned in section 2.1 (and section 1.4.1 of the previous chapter), if quantificational determiners are devices for relating predicates, a question arises: Why are

---

[9] An even stronger way to show that *every* is not understood as a relation would be to find evidence that adopting a two-group comparison strategy would, in fact, lead to superior performance, even at evaluating the initial quantificational statement. Pietroski et al. (2009) ran an analogous control experiment, discussed in Chapter 1. There, participants avoided a one-to-one correspondence strategy when asked to evaluate *most*-sentences with respect to dot-displays; but if asked to *find the leftover dot* with respect to the same displays, they were more accurate. Given that the one-to-one strategy was available and superior, it is striking that participants avoided it.

only certain relations lexicalized? Natural languages make use of only a small corner of the space of possible two-place relations. If *every F is G* expresses the relation "$F \subseteq G$", why does no natural language determiner express the converse relation "$F \supseteq G$"?

Of particular interest since Barwise and Cooper (1981), Higginbotham and May (1981), and Keenan and Stavi (1986) has been the robust generalization that all natural language determiners are "conservative," described in section 2.3.1. The relational conception of determiner meanings allows for one way of stating the generalization – all determiner meanings R are such that $R(X, Y) \equiv R(X \cap Y, Y)$ – but it makes the generalization difficult to explain. On the other hand, while the "conservativity" universal is mysterious given a relational view, it is a logical consequence of restricted quantification. So while there are ways of retaining relationality and capturing conservativity (sections 2.3.2 and 2.3.3), abandoning relationality altogether offers a simple explanation of this cross-linguistic universal (section 2.3.4).

2.3.1 The "conservativity" universal

The central observation is this: For any determiner, repeating its internal argument within its external argument is logically insignificant. For example, the meanings of (94a) and (94b) are logically equivalent (i.e., (94a) and (94b) pattern together with respect to truth/falsity).

(94)   a. Every frog is green

        b. Every frog is a frog that is green

More generally, an instance of (95a), where PRED can be a verbal or adjectival predicate, is logically equivalent to the corresponding instance of (95b), allowing variations in morphology and word order. If (95a) is TRUE, (95b) will be TRUE, and vice versa.

(95)   a. [[DET NP] PRED]

        b. [[DET NP] [be NP that PRED]]

This phenomenon has been described in various ways, but is often called the "conservativity" constraint. Importantly, all natural language determiners seem to obey this constraint.[10]

That all determiners follow the "conservativity" constraint is surprising, as it's easy to imagine hypothetical determiners that fail to be "conservative." For example, suppose that *equi* meant "equal in number" so that (96a) means that there are the the same number of frogs and green things, and (96b) means that there are the same number of frogs as green frogs.

(96)   a. Equi frogs are green

            ≈ *the frogs and the green things are equal in number*

        b. Equi frogs are frogs that are green

            ≈ *the frogs and the green frogs are equal in number*

These two sentences are not logically equivalent. Imagine a domain consisting of

---

[10] Reverse proportional readings of sentences like *many Scandinavians have won the Nobel prize in literature* have been used to argue that *many* is a possible counter example to this generalization (Westerståhl, 1985). But Romero (2015) argues that these readings result from combining a conservative determiner with degree operator.

one green frog, one blue frog, and one green apple. This situation could be accurately described by (96a) (2 frogs = 2 green things) but not by (96b) (2 frogs ≠ 1 green frog). Taking the domain to be the actual world, (96a) seems obviously FALSE, since there are more green things than there are frogs; but (96b) would be TRUE if not for the existence of non-green frogs. In any case, the fact that that these two sentences fail to be logically equivalent means that *equi* does not have a "conservative" meaning. And, in line with the generalization, no natural language has a determiner that means "equal in number."

To take another example of an imaginary "non-conservative" determiner, suppose *yreve* meant "includes" so that (97a) means that the frogs include all green things, and (97b) means that the frogs include all the green frogs.

(97) a. Yreve frogs are green

 ≈ *the frogs include the green things*

 b. Yreve frogs are frogs that are green

 ≈ *the frogs include the green frogs*

To a first approximation, this is the meaning that *only* would have if it were a determiner (though as discussed in section 2.2.5, *only* is a focus operator). Like the *equi*-sentences, these two *yreve*-sentences are not logically equivalent. Imagine a domain with a green frog and a green apple; (97a) cannot be used to describe such a situation (because there is a green non-frog), though (97b) can (in fact, regardless of the domain, (97b) will always be TRUE).

Both of these hypothetical determiners have seemingly simple meanings and

would potentially be communicatively useful, so it's striking that English and other languages lack them. This does not appear to be a historical accident either. Hunter and Lidz (2013) were able to teach 5-year-olds a novel quantifier when it had a conservative but cross-linguistically unattested meaning, like *not all*. But they were unable to teach 5-year-olds a novel quantifier in the same experimental set-up if the intended meaning was the non-conservative *not only*.

In particular, children were asked to help sort scenes depicting boys and girls on the beach and at the park based on the preferences of a picky puppet. The experimenter told children that the puppet likes it when (98), and that it was their job to figure out what *gleeb* means.

(98) Gleeb girls are on the beach

    a. ≈ *not all of the girls are on the beach*

    b. ≈ *not only girls are on the beach*

When the puppet liked cards in which (98a) was TRUE, 5-year-olds were able to correctly sort new scenes according to whether gleeb girls were on the beach after seeing just five examples. But when the puppet liked cards in which (98b) was TRUE, participants trained and tested using the same procedure were at chance performance on novel scenes (though see Spenader and de Villiers (2019), who found no learning in either condition in a similar task).

In this task at least, learners did not seem to consider the non-conservative "not only" as a possible determiner meaning. This impressive finding awaits empirical replication. But to the extent that it generalizes, it suggests that conservativity is a

deep fact about the language faculty: Non-conservative determiners are conceivable but not possible lexical items for humans. Semantic theory should aim to explain this robust and seemingly deep cross-linguistic constraint.

2.3.2 Retaining relationality: Lexical filtering

Holding fixed the idea that quantificational determiners express relations, the task is to explain why only a small fraction of the possible relations are available for lexicaliztation. If *every F is G* means "$F \subseteq G$", then the hypothetical *equi F is G* (in (96a)) means "$F = G$" and the hypothetical *yreve F is G* (in (97a)) means "$F \supseteq G$". What rules out relations like "$=$" and "$\supseteq$", which prima facie seem conceptually related to "$\subseteq$"?

One straightforward approach would be to forbid the problematic relations at the lexical level. Keenan and Stavi (1986) offer one version of this approach. First, they define a store of "basic" relations that are conservative (including the relations purportedly expressed by *every, at least four/at most three*, and *a/some $NP_{singular}$/at least 1*). Then, they combine this set of relations with a set of conservativity-preserving operators such that more complex relations (e.g., the one expressed by *most* or the one expressed by *seventeen*) can be constructed by combining the primitives with the operators. If these are the only relations available for lexicaliztation, then only conservative relations can be lexicalized.

This is not to say that relations like identity or equality are absent from cognition, just that "$X = Y$" and "$|X| = |Y|$" are not among the primitives available for

lexicaliztation and cannot be constructed out of the primitives and operations that are available. In a sense then, this way of accounting for the phenomena amounts to a redescription of the facts in terms of the idea that determiners express relations.

Moreover, an approach that "filters" out the problematic relations at the lexical level does not itself explain the asymmetry between the determiner's two arguments. If determiners are tools for expressing relations, then the internal and external argument of a determiner have the same logical status: Both are just terms in a relation. Why, then, does *every F is G* mean "$F \subseteq G$" and not "$G \subseteq F$"?

We might suspect the blame should be placed on the syntax of the expression. After all, in a sentence like *every frog is green* in (99), *every* combines with its internal argument (*frog*) first and (assuming quantifier raising) its external argument (*t is green*) contains a trace of displacement.

(99)

```
                    S
           _____/_____
          DP                VP
        __/\__           __/\____
       D    NP          t is green
       |     |
       |     |
     Every  frog
```
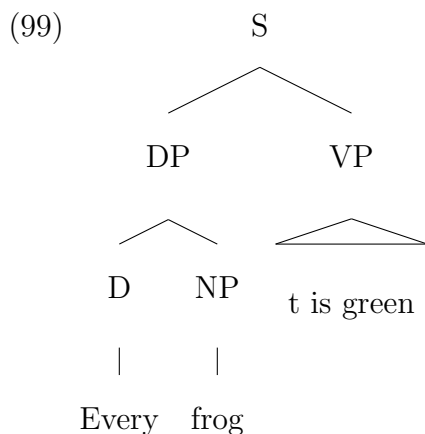
But the relational view obscures this distinction, treating both the internal and the external argument as names for groups that are related by the determiner (the frogs and the green things). As a result, it is only an accident that the grammatically internal argument ends up occupying the position on the lefthand side of the rela-

116

tion. So while the "lexical filtering" approach may ultimately be right, it is more stipulative than we might like.

2.3.3 Retaining relationality: Interface filtering

In an attempt to embrace the syntactic distinction between internal and external arguments but preserve the relational conception of determiner meanings, Romoli (2015) offers an alternative (that builds off of suggestions from Chierchia (1995), Fox (2002), Ludlow (2002), and Sportiche (2005)). Essentially, the idea is that a sentence pronounced as *every frog is green* actually means something more like *every frog is a frog that is green*.

More formally, the idea relies on the ancillary hypothesis that in a sentence like (99), the trace is interpreted as a second instance of the noun phrase *frog*. The displaced determiner is not interpreted in the same way, but is converted into an instance of *the*. For example, (100a) ends up with the logical form in (100b).

(100)   a. [Every frog [~~every frog~~ is green]]

   b. [Every frog [~~the frog~~ is green]]

   c. $\{x : frog(x)\} \subseteq \{x : frog(x) \,\&\, green(x)\}$

If the second instance of *frog* (the one in the external argument) is interpreted conjunctively, then (given a relational meaning for *every*), (100a) has the meaning in (100c). On this view, *every* can still be said to express the same relation as is traditionally assumed ("$\subseteq$"), but instead of relating the set of frogs and the set of

green things, it relates the set frogs and the set of green frogs.[11] This leaves the truth-conditions unchanged (if this were not the case, then "⊆" would fail to be a conservative relation). It is worth noting, however, that the experiments reported in section 2.2 tell against this hypothesis, as the group defined by the conjunction of both arguments is never explicitly represented by participants.

In any case, given this way of interpreting traces, quantificational sentences effectively end up with a second instance of the internal argument in the external argument. Consider what this would mean for a non-conservative determiner like *equi* (i.e., the equal-in-number relation) from (96a). A sentence like (101a) would end up with the logical form in (101b) and the interpretation in (101c).

(101) a. [Equi frog [~~equi frog~~ is green]]

  b. [Equi frog [~~the frog~~ is green]]

  c. $\{x : frog(x)\} = \{x : frog(x) \ \& \ green(x)\}$

But (101c) is truth-conditionally equivalent to (100c). So when used in a sentence, *equi* ends up being truth-conditionally indistinguishable from *every*. In fact, Romoli (2015) notes that this leads to the possibility that *every* might in fact express the "=" relation, not the "⊆" relation (see note 11). In any case, the non-conservative "the frogs and the green things are equal in number" interpretation never arises.

---

[11] While *every* can be said to express "⊆", this is somewhat redundant since "=" does all the work and "⊂" never comes into play. That is, while it's impossible for the frogs to be a proper subset of the green frogs, *every frog is green* obtains when the frogs are identical to the green frogs. So if the external argument of *every* really is a conjunction of what we thought was the external argument and the internal argument, it seems more parsimonious to say that *every* expresses the identity relation. Given that "=" is not conservative, this would be an unexpected conclusion.

This reasoning holds for a subset of the would-be-non-conservative determiners: Given the interpretation of traces discussed above, sentences with such determiners would end up being truth-conditionally equivalent to sentences with a conservative determiner. So, contra the "lexical filtering" view, lexical items that express non-conservative relations might exist, but, in practice, they never lead to non-conservative sentence meanings.

This reasoning does not apply to all potential non-conservative determiners, however. Consider another class of problematic hypothetical quantifiers, that includes *yreve* (i.e., *only* as a determiner) from (97a). After trace conversion, the resulting interpretation – given in (102c) – is not truth-conditionally equivalent to that of any conservative determiner.

(102)   a. [Yreve frog [~~yreve frog~~ is green]]

      b. [Yreve frog [~~the frog~~ is green]]

      c. $\{x : frog(x)\} \supseteq \{x : frog(x) \ \& \ green(x)\}$

In fact, (102c) will always be TRUE, since the set of frogs must be a superset of or be identical to the set of green frogs.

To rule out *yreve*, a filter on trivial sentence meanings – sentences that are either tautologies or contradictions, given any substitution of lexical content – is posited (Fox and Hackl, 2006; Gajewski, 2002). If *yreve* existed, any sentence with it would be declared ungrammatical given such a filter. In that way, Romoli (2015)'s approach "filters out" problematic determiners thanks to details of the syntax-semantics interface.

There are other hypothetical non-conservative determiners about which this "interface filtering" view is silent. Consider a determiner *onemore* that means "outnumbers by one." So (103a) means that the frogs outnumber the green things by one and (103b) means that the frogs outnumber the green frogs by one.

(103)   a. Onemore frog is green

    ≈ *the frogs outnumber the green things by 1*

    b. Onemore frog is a frog that is green

    ≈ *the frogs outnumber the green frogs by 1*

Onemore is not conservative, since we can imagine a situation where (103a) is TRUE but (103b) is not. For example, if there were two blue frogs and a green apple, the frogs would outnumber the green things by one (2 vs. 1) but the frogs would outnumber the green frogs by two (2 vs. 0). So *onemore* must be ruled out as a possible determiner.

Given the interpretation of traces discussed above, (104a) would have the meaning in (104c).

(104)   a. [Onemore frog [~~onemore frog~~ is green]]

    b. [Onemore frog [~~the frog~~ is green]]

    c. $|\{x : frog(x)\}| - 1 = |\{x : frog(x)\ \&\ green(x)\}|$

The meaning in (104c) is not truth-conditionally equivalent to any existing conservative determiner, but it also is not trivial. The example above shows that it can be FALSE (2 blue frogs and 0 green frogs). But it can also be TRUE: In a context

consisting of one blue frog and one green frog, the frogs outnumber the green frogs by one (2 vs. 1). So *onemore* is not conservative but fails to be ruled out by either of the two approaches posited by the "interface filtering" view.

To be sure, one could argue that this kind of non-conservative determiner is ruled out on other grounds. For example, *onemore* runs afoul of another generalization: Morphologically simple determiners are monotonic (Barwise and Cooper, 1981). That is, either a subset → superset or a superset → subset inference is licensed in each argument. Given *every frog is green*, we can infer *every big frog is green* (superset → subset) and *every frog has a color* (subset → superset). But from the fact that the frogs outnumber the green things by 1 (*onemore frog is green*), it follows neither that the big frogs outnumber the green things by 1 (imagine 2 small blue frogs and 1 big green frog) nor that the amphibians outnumber the green things by 1 (imagine 2 blue frogs and 1 green tadpole). So if the hypothetical *onemore* is morphologically simple, it might be ruled out by failing to be monotonic.

But morphologically complex determiners can fail to monotonic. For example, consider *all but one*. From *all but one frog is green*, we can infer neither *all but one big frog is green* (imagine 2 small green frogs, 1 small blue frog, and 2 big green frogs) nor *all but one amphibian is green* (imagine 2 small green frogs, 1 small blue frog, and 1 blue tadpole). So if *onemore* were morphologically complex, it could not be ruled out on grounds of being non-monotonic.

We might, then, be tempted to say that *onemore* is ruled out on the basis of communicative uselessness. Perhaps it's just not the sort of thing people would ever want or need to say. It could get lexicalized, but it turns out that it doesn't.

But this begins to make conservativity look less like a deep fact about the language faculty and more like a conspiracy of various independent factors.

2.3.4 Abandoning relationality: Ordered predication

In contrast, the non-existence of non-conservative determiners is a logical consequence of restricted quantification. This intuition has been cashed out in various ways: By Pietroski (2005, 2018) in a framework where the basic semantic operation is conjunction, and meanings instructions to retrieve and assemble concepts; by Westerståhl (2019) in set-theoretic terms; and by Lasersohn (2021) in a framework that treats common nouns like *frog* as restricted variables instead of as predicates. The crucial point is that all conservative determiners can be stated as restricted quantifiers, but non-conservative determiners cannot.

The reason is that restricted quantification enforces what we might call "ordered predication." Namely, the internal predicate first restricts the domain, then the external predicate supplies a further condition that gets predicated of the restricted domain. The determiner itself specifies how that further condition should apply. This creates no problems for conservative determiners, as all of them can be stated relative to a restricted domain. To see this, consider the following paraphrases of *every frog is green*, *most frogs are green*, and *no frogs are green*, using the "hook" notation from section 2.1 to signify relativization:

(105)   a. $\text{EVERY}_x(green_x) \upharpoonright \underline{frog_x}$

         $\approx$ *Relative to the frogs, everything is green*

b. $\textsc{most}_x(green_x) \upharpoonright \underline{frog_x}$

   $\approx$ *Relative to the frogs, most things are green*

c. $\textsc{no}_x(green_x) \upharpoonright \underline{frog_x}$

   $\approx$ *Relative to the frogs, no things are green*

Any conservative determiner can be relativized in this way. But non-conservative determiners cannot. For example, recall that *yreve* is a hypothetical non-conservative determiner that has the meaning that *only* would have if it were a determiner. So, abstracting away from focus sensitivity, *yreve frogs are green* means the same thing as *only frogs are green*. But as (106) should make clear, this meaning cannot be paraphrased in the same way.

(106) $\textsc{only}_x(green_x) \upharpoonright \underline{frog_x}$

   $\approx$ *Relative to the frogs, only things are green*

The problem is that the meaning of *yreve* requires making reference to the external argument independently of the internal argument. It cannot be stated in terms of how the predicate *is green* applies to the frogs. The same is true for other non-conservative determiners. In *equi frogs are green* – which has the intended meaning that the frogs and the green things are equal in number – the cardinality of the green things, independent of the frogs, needs to be taken into account. If the domain is restricted to just the frogs, this cardinality is inaccessible.

In sum, restricted quantification respects the syntax by maintaining that the two predicates apply in a certain order and play different logical roles. That restricted quantifiers are conservative follows as a logical consequence.

2.4 Chapter summary

This chapter provided linguistic and psycholinguistic reasons for thinking that *every* and *all* have semantic representations that are restricted, like (107). This stands in contrast to the standard view that determiners are special cases of generalized quantifiers and have semantic representations that are relational, like (108).

(107) $\lambda\mathbf{I}.\lambda\mathbf{E}.\iota G(\forall x[Gx \equiv \mathbf{I}x])[\mathbf{E}(G)]$

(108) $\lambda\mathbf{I}.\lambda\mathbf{E}.\iota G(\forall x[Gx \equiv \mathbf{I}x])\{\iota G'(\forall y[G'y \equiv \mathbf{E}y])[G \subseteq G']\}$

The main difference, discussed in section 2.1, is in the logical role played by the two arguments. In restricted representations, the internal and external arguments serve distinct logical roles. The internal argument restricts the domain in some way, whereas the external argument provides an additional predication. On the relational view, both serve identical roles: terms in a relation.

In support of the restricted view, section 2.2 presented seven experiments showing that participants asked to evaluate sentences like *every big circle is blue* explicitly represent the group named by the internal argument, but not the group named by the external argument. Avoiding a relate-two-groups verification strategy makes sense if *every* has a restricted representation like (107) but would be hard to explain if it had a relational representation like (108). Participants also avoided representing the group named by the conjunction and the internal and external arguments, which supports the single-group restricted representation (107), as opposed to some other restricted representation.

On the semantic side, section 2.3 discussed the robust cross-linguistic universal that all determiners are "conservative." The relational view makes this generalization somewhat mysterious, as many relations fail to be "conservative" and must be ruled out as possible meanings in order to explain the cross-linguistic constraint. Approaches that aim to explain conservativity while retaining a relational conception of determiner meanings were considered in sections 2.3.2 and 2.3.3. But abandoning relationality provides a simple explanation: No determiners in natural language express "non-conservative" relations because no determiners express relations in the first place.

Chapter 3: First-order vs. second-order quantification

Chapter 2 argued that quantificational determiners are restricted, not relational. But there are many ways to be a restricted quantifier. This chapter will provide some reasons for thinking that *each* has a first-order representation whereas *every* and *all* have second-order representations. Though it builds off of the results of the previous chapter, the first-order/second-order distinction is independent of the restricted quantification hypothesis.

Section 3.1 describes the logical distinction in greater detail and shows that the proposed representations for the universal quantifiers align with the fact that *each* differs from the other universals in being strongly distributive. Section 3.2 discusses independently-supported psychological systems that parallel the logical first-order/second-order distinction. Section 3.3 then leverages known properties of these representations in a series of experiments. To preview the results: Adults and children encode group-properties (cardinality and center of mass) of the first-argument when evaluating *every*-sentences. But when *every* is replaced with *each*, they instead encode individual-properties (color). Section 3.4 argues that the proposed representations offer a novel explanation for the observation that *every* is compatible with certain kinds of generic interpretations that *each* resists.

3.1 The logical distinction

The first-order/second-order distinction is, in the first place, a distinction in logical syntax. The difference is in which variable position gets quantified into. First-order representations only quantify into object (i.e., lowercase) variable positions whereas second-order representations can also quantify into predicate (i.e., uppercase) variable positions. By way of illustration, suppose that (109a) represents the proposition *kermit is green.* If we take (109a) as a premise, we can make the first-order inference in (109b) that *there is something that is green* and the second-order inference in (109c) that *there is some property that applies to kermit.*

(109)   a. Gk

   b. $\exists x(Gx)$ (first-order inference)

   c. $\exists X(Xk)$ (second-order inference)

In (109b), the first-order quantifier "$\exists x$" binds the first-order variable x in "Gx". In (109c), the second-order quantifier "$\exists X$" binds the second-order variable X in "Xk".

To be sure, there are different views about how to interpret second-order variables. For Frege (1879, 1893), a predicate like *is green* corresponds to the function that maps all and only green things to truth. So (109c) might be glossed *there exists some function that maps kermit to truth.* In set-theoretic terms, (109c) might be glossed *there exists some set such that kermit is an element of it.* Alternatively, Boolos (1984) argues for treating second-order variables as plural variables that range

over the same values as first-order variables. They differ only in that they can take on multiple such values at once. On this view, (109c) would be glossed *there exist some things such that kermit is one of them.*[1]

The fact that there are different interpretations of second-order variables is worth dwelling on. Second-order logic is often interpreted set-theoretically. But we can separate the claim that a representation is second-order from the claim that a representation implicates a set (or any other plural entity, like a mereological sum). Given the plural logic way of interpreting second-order variables described above, a representation containing a second-order quantifier need not imply any additional ontological commitments beyond those implied by a first-order representation. So, (109c) does not necessarily imply that the domain includes a set X.

Likewise, on the plural logic understanding, the second-order (110) does not imply that the domain includes the set of green things.

(110) $\exists X(\forall y[Xy \equiv Green(y)])$

$\approx$ *There are some things$_X$ such that each thing$_y$ is one of them$_X$ iff it$_y$ is green*

But, by virtue of it being second-order, (110) does allow for grouping the green things in the sense that all of them can be considered at once. That is, instead of

---

[1] One of Boolos' main reasons for proposing a plural logic understanding of second-order variables is that it avoids Russell's paradox (for more discussion on this point, see Schein (1993), Pietroski (2003a), and Pietroski (2005)). If the expression $\exists X\forall y[Xy \equiv y \notin y]$ is understood to mean *there is a set X whose members are all and only the things that are not members of themselves*, paradox ensues. Namely, is the set X a member of itself? If it is self-elemental, then it cannot be an element of X by definition, meaning it isn't self-elemental. But if it isn't self-elemental, then it is by definition an element of X, meaning it is self-elemental. For a "real-world" example, consider a barber X who shaves all and only the barbers who don't shave themselves. Does X shave themself? As Boolos pointed out, no paradox arises if $\exists X$ does not imply a separate object X. If the expression in question is understood as *there are some things$_X$ that have the following property: Each thing$_y$ is one of them$_X$ iff it$_y$ is non-self-elemental*, then it is simply true, since there are plenty of things that aren't elements of themselves (e.g., me, you, this dissertation).

saying that the variable X has one value – the set of green things – we can say that this variable has multiple values: all of the individual green things.

The dissociation between sets and second-order quantification runs in the other direction as well: Implicating a set does not make a representation second-order. Suppose we explicitly include the set of green things in our domain. We can still describe *kermit is a member of the set of green things* in a completely first-order way, as in (111), which doesn't rely on representing the elements of the set collectively.

(111) $\exists s(s = \{x : x \ is \ green\} \ \& \ kermit \in s)$

    $\approx$ *There exists a set, identical to the set of green things, and kermit is an element of it*

Even some relations between sets can be specified in completely first-order terms. Suppose we assume that *frog* denotes the set of frogs and *is green* denotes the set of green things. We can describe *every frog is green* with restricted quantification as in (112), which doesn't rely on representing the elements of either set collectively.

(112) $\forall x(x \in \{x : x \ is \ a \ frog\})[x \in \{x : x \ is \ green\}]$

    $\approx$ *each thing that is a member of the set of frogs is such that it is a member of the set of green things*

In sum, at issue with the first-order/second-order distinction is what type of variable the quantifier targets. First-order representations contain only quantifiers that bind lowercase variables. Second-order representations allow for binding uppercase variables as well. Whether quantification into a second-order variable

position implicates the existence of plural entities like sets depends on how upper-case variables are interpreted. To remain agnostic on this point, we can say that a representation implicates a "group". This is not intended to be understood as a technical term, for example, in the sense of Landman (1989a,b) (for whom groups are entities that are quite detached from their members). Instead, "group" is intended to be neutral between two understandings: "some sort of plural entity, like a set/sum/aggregate/lattice" or "a way of grouping the relevant predicate in the Boolos (1984) sense."

3.1.1 First-order and second-order universal quantifiers

As discussed in the preceding chapters, quantificational determiners like *each* and *every* combine with two arguments. The first – the internal argument $\mathbf{I}$ – is nominal, like *frog* in (113) and (114). Assuming quantifier raising, the second – the external argument $\mathbf{E}$ – is sentential, like *t is green* in (113) or *kermit saw t* in (114).

(113) Every frog is green

(114) Kermit saw every frog
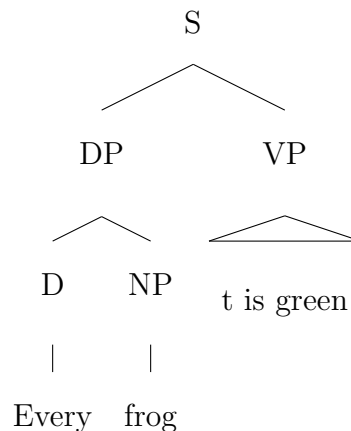
```
                              S
                         /        \
                      DP           VP
                     /  \         /    \
                    D    NP    Kermit saw t
                    |    |
                  Every  frog
```
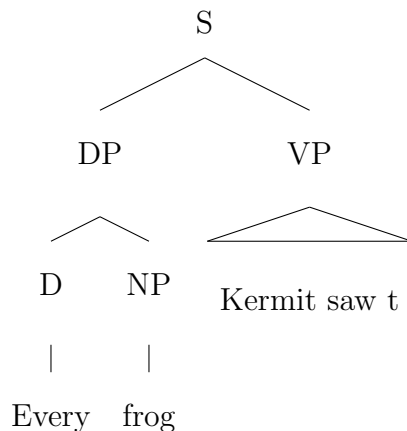
Both the internal and the external argument name predicates and, in some sense, the determiner relates these two predicates. The question at issue here is how to specify the determiner's contribution. We can adopt the standard assumption that the determiner *every* is specified in terms of abstractions from the meanings of the two predicates with which it combines, as in (115).

(115) $\lambda\mathbf{I}.\lambda\mathbf{E}.\text{EVERY}(\mathbf{I}, \mathbf{E})$

In a sense, (115) is a second-order representation, because the lambda expressions bind uppercase variable positions. However, as discussed in Chapter 1, "$\lambda\mathbf{I}$" and "$\lambda\mathbf{E}$" for present purposes are meant as technical devices, not psychological hypotheses.[2] The psychological question is how to specify what occurs on the right side of the lambda expressions, "$\text{EVERY}(\mathbf{I}, \mathbf{E})$". Consider again the four hypotheses introduced in Chapter 1, (116) - (119).

---

[2] We could instead give a syncategorematic treatment of *every* that avoids lambda expressions. On this way of doing things, the entire specification could be taken as a psychological hypothesis. Or, we could treat the lambda expressions as part of the psychological hypothesis and consider alternative hypotheses about how composition proceeds. The current proposal is agnostic about these questions.

(116) $\forall x(\mathbf{I}x)[\mathbf{E}x]$ $\hspace{4cm}$ (FO restricted, 0G)

(117) $\iota G(\forall x[Gx \equiv \mathbf{I}x])[\mathbf{E}(G)]$ $\hspace{4cm}$ (SO restricted, 1G)

(118) $\iota G(\forall x[Gx \equiv \mathbf{I}x])\{\iota G'(\forall y[G'y \equiv Gy \ \& \ \mathbf{E}y])[G = G']\}$ $\hspace{1cm}$ (SO restricted, 2G)

(119) $\iota G(\forall x[Gx \equiv \mathbf{I}x])\{\iota G'(\forall y[G'y \equiv \mathbf{E}y])[G \subseteq G']\}$ $\hspace{1.5cm}$ (SO relational, 2G)

The representation in (116) – *for each thing$_x$ such that it$_x$ is $\mathbf{I}$, it$_x$ is $\mathbf{E}$* – is completely first-order, since only a lowercase variable (x) is quantified into. In contrast, (117)-(119) are second-order since they each contain "ιG", which binds the second-order ("group") variable G. This variable groups together the satisfiers of the internal predicate $\mathbf{I}$. As discussed above, there are multiple ways of interpreting the second-order quantifier "$\iota G(\forall x[Gx \equiv \mathbf{I}x])$". If assignments of values to variables allow only one value to be assigned to G, then G ranges over plural entities of some sort (e.g., sets) and it can be read as (120a). If multiple values are allowed, then G ranges over the same entities as first-order variables and it can be read as (120b).

(120) $\iota G(\forall x[Gx \equiv \mathbf{I}x])[...G...]$

       a. $\approx$ *the set of things that meet the $\mathbf{I}$ condition is such that it...*

       b. $\approx$ *the one or more things that meet the $\mathbf{I}$ condition are such that they...*

On the former view, "$\mathbf{I}x$" is shorthand for set inclusion: $x \in \{x : \mathbf{I}x\}$. But on the latter view, "$\mathbf{I}x$" is shorthand for *x is one of the $\mathbf{I}$s*.

In any case, in (117), "$\iota G(\forall x[Gx \equiv \mathbf{I}x])$" is the only second-order quantifier, where G is the group that meets the $\mathbf{I}$ condition. The universality is specified

by saying that this group G meets the **E** condition: "$[\mathbf{E}(G)]$" (e.g., *the frogs are green*).[3] In (118), there are two second-order variables quantified into: the group that meets the **I** condition, G, and the group that meets both the **I** condition and the **E** condition, G'. The universality is specified by saying that these groups are identical: "$[G = G']$" (e.g., *the frogs are the green frogs*). Likewise, in (119), there are two second-order variables quantified into. As before, G is the group that meets the **I** condition, but in this case, G' is the independent group that meets the **E** condition. The universality is specified by saying that the former is a subset of/is among the latter: "$[G \subseteq G']$" (e.g., *the frogs are among the green things*).

Because (119) groups and relates the satisfiers of both predicates independently, it is relational in the sense of Chapter 2. In contrast, (116)-(118) are all restricted. This is clearest for (116), as only things that meet the **I** condition are considered with respect to the **E** condition. The same is true in (117), save for the difference that the things that meet the **I** condition are grouped under the variable G. Things are less clear with (118), which implicates two groups – G and G' – and relates them. This specification is nonetheless restricted because G' is not independent of G. Rather, G' is an additional selection made against a restricted domain that only includes G.

In any case, the experiments reported in Chapter 2 tell against *every* having a representation that implicates two groups, like the restricted (118) or the

---

[3] What exactly it means for a set of things or for one or more things to meet the **E** condition depends on what the **E** condition is. In general, since **E** will be an open sentence (e.g., *t is green*), we can think of G as filling the open slot left after displacement (e.g., *G is green*). But when the predicate in question is lexically distributive (like *be green*), each individual must meet the condition; see section 3.1.2. Thanks to Roman Feiman for helpful discussion on this point.

unrestricted (i.e., relational) (119).[4] Building on those findings, this chapter will argue that *every* has a representation that implicates one group, like the second-order (117), whereas *each* has a representation that implicates zero groups, like the first-order (116).

3.1.2   Dealing with distributivity

Before coming to the semantic and psychosemantic arguments for the proposed representations, it is worth considering how these representations interact with distributivity. Accounting for data related distributivity is not the main reason for thinking that some determiners have first-order representations and others have second-order representations. The main reason is the proposed interface with known psychological systems, discussed starting with section 3.2. But distributivity is perhaps the most-discussed topic surrounding the universal quantifiers so it is important that the proposed distinction is consistent with the basic facts in this domain.

A common theme in the literature since Vendler (1962) is that *each* forces distributivity: The predicate in question must apply to each individual, not to the group as a whole. This makes *each* incompatible with certain collective predicates like *gather* and *surround*, which necessarily apply to the entire group. At the same time, *all* and (arguably) *every* can combine with these predicates, as in (121).

---

[4] If being a device for creating restricted quantifiers is a universal property of determiners, then (119) should not be a possible determiner meaning. But (118) remains a plausible hypothesis. It may turn out to be the meaning of *all* (though the experiment in section 2.2.4 of the previous chapter provides some initial reason to doubt this possibility). For simplicity, this chapter will often group *every* and *all* together and will focus more on *every*, as it constitutes a true minimal pair with *each*.

(121)   a. All (of the) frogs gathered in the forest

         b. Every frog gathered in the forest

         c. #Each frog gathered in the forest

To be sure, (121a) is preferred over (121b) for expressing the relevant thought. Perhaps as a result of this relative preference, many authors go so far to report (121b) with a #, ?, or even * (e.g., Gil 1992; Beghelli and Stowell 1997; Tunstall 1998; Winter 2002; Champollion 2020). This seems wrong. At the least, the acceptability of (121b) is an open question that should be subject to a controlled acceptability judgement experiment. That said, some collective predicates, like (122), do seem to be incompatible with *every* despite being compatible with *all*.

(122)   a. All the elms clustered in the forest

         b. ?Every elm clustered in the forest

And other collective predicates, like those in (123), are bad with any universal quantifier, including *all* (Dowty 1987; Winter 2002; Champollion 2020).

(123) #All the frogs in the forest {are numerous/are a good team}

Whatever the status of *every* and *all* and certain collective predicates though, it is clear that *each* always gives rise to distributive interpretations whereas *every* and *all* do not. To take an example with an ambiguous predicate, in (124a) there were as many piano-lifting events as there were students. But while (124a) cannot describe a situation where the students worked together to collectively accomplish the task, it seems that (124b) and (124c) can.

(124)   a. Each student lifted the piano (#together/by themselves)

        b. Every student lifted the piano (together/by themselves)

        c. All (of the) students lifted the piano (together/by themselves)

Again, we should be careful to separate preference from compatibility. Clearly (124c) is preferred for expressing the collective thought in which the piano got lifted one time by the whole group. And (124a) is preferred for expressing the distributive thought in which the individual students each got a turn. This calls for explanation. But at issue here is the asymmetry: (124b) and (124c) have distributive and collective interpretations, whereas (124a) has no collective interpretation.

If a second-order representation underlies *every* and *all*, it needs to be compatible with both distributive and collective interpretations. Likewise, if a first-order representation underlies *each*, it needs to be incompatible with collective interpretations and collective predicates. The remainder of this section aims to show that the proposed representations in (116) and (117) meet these requirements.

3.1.2.1   Manifestly distributive or collective predicates

Predicates like *smile*, *blink*, *take a deep breath*, *see the photo*, and *be green* are always interpreted distributively. As noted above, this is true regardless of the subject (e.g., *each student smiled* is no more distributive than *all the students smiled* or the explicitly "group-denoting" *the whole class smiled*). These manifestly distributive predicates are often referred to as cases of L(exical)-distributivity (e.g., Winter

1997; de Vries 2015).[5] L-distributivity can be captured by building distributivity directly into the lexical specification of the predicate, or with an associated meaning postulate (e.g., Hoeksema 1983; Scha 1984; Roberts 1987; Champollion 2017). For example, we might stipulate that what it means for *smile* to apply to a group of individuals X is for it to apply to each one of them, as in (125).

(125) $Smile(X) \equiv \forall y(Xy)[Smile(y)]$

More needs to be said about which predicates come with meaning postulates like (125) and why (see Glass (2021) for a survey of distributivity inferences arising from a variety of predicates). But regardless, meaning postulates of this sort are often invoked to deal with L-distributivity. Formally, we can say that the **E** condition is normally understood to apply to the group, as in (126). But when the **E** condition is an L-distributive predicate like *smile* or *blink*, "**E**(G)" is understood distributively, as in (127).

(126) $\lambda\mathbf{I}.\lambda\mathbf{E}.\iota G(\forall x[Gx \equiv \mathbf{I}x])[\mathbf{E}(G)]$ =(117)

(127) $\lambda\mathbf{I}.\lambda\mathbf{E}.\iota G(\forall x[Gx \equiv \mathbf{I}x])[\mathbf{E}(G) \equiv \forall y(Gy)[\mathbf{E}y]]$ (if **E** is L-distributive)

This yields the desired representations for sentences like *each/every/all student(s) smiled* in (128). The first-order case in (128a) can be glossed as *for each thing such that it is a student, it smiled.* And the second-order case in (128b) can be glossed as *the things that are students are such that each one of them smiled.* In either case, the predicate is forced to distribute down to the individual students.

---

[5] Lexical-distributivity is sometimes also called Predicate-distributivity, since the distributivity is generated in virtue of what the predicate means and without the need for a covert operator. This is meant to stand in contrast with Phrasal-distributivity (sometimes also called Quantificational-distributivity), which is argued to require a covert quantificational operator.

(128)   a. $\forall x(\mathbf{Student}(x))[\mathbf{Smile}(x)]$ (*each*)

b. $\iota G(\forall x[Gx \equiv \mathbf{Student}(x)])[\mathbf{Smile}(G) \equiv \forall y(Gy)[\mathbf{Smile}(y)]]$ (*every/all*)

Turning to the manifestly collective predicates, like *gather*, the only difference is that there is no distributive meaning postulate present. Everything else remains the same. In the first-order case in (129a), the predicate is always forced to distribute down to the individuals, since there is no device for grouping them. This leads to an anomaly, because the predicate *gather* must apply to multiple individuals at once (i.e., there is no distributive way to gather). The second-order (121b), on the other hand, allows for the predicate to apply collectively (i.e., *the things that are students are such that they gathered*).

(129)   a. $\forall x(\mathbf{Student}(x))[\mathbf{Gather}(x)]$ (#*each*)

b. $\iota G(\forall x[Gx \equiv \mathbf{Student}(x)])[\mathbf{Gather}(G)]$ (*every/all*)

In sum, the first-order/second-order distinction creates no problems for capturing the basic data regarding manifestly distributive and manifestly collective predicates. A first-order representation is not compatible with collective predicates, as the condition is only applied to individuals. This is in line with the fact that *each* is not compatible with collective predicates. A second-order representation is compatible with both distributive and collective predicates, though the former requires invoking a meaning postulate associated with the predicate. This is in line with the data regarding *every* and *all*.

3.1.2.2   Ambiguous predicates

As noted above, predicates like *lifted the piano* or *sang happy birthday* are ambiguous between collective and distributive readings (or, at least, underspecified as to which reading is intended). Neither reading entails the other. It can be true that it took all the students to get the song sung but false that each student sang happy birthday themselves (e.g., if they took turns and each sang only one line). Conversely, it can be true that each student sang happy birthday individually without it being true that there was a single event of singing in which all the students collectively participated (e.g., in the context of auditioning for the school musical).

Ambiguous predicates like these are often referred to as P(hrasal)-distributivity. Unlike L-distributivity, P-distributivity is usually argued to require more than a meaning postulate (Champollion, 2020; Landman, 2012; Winter, 2000). This is because P-distributivity is exemplified by predicates like *formed a circle*, in which case it is the whole verb phrase, not just the verb, that optionally has a distributive interpretation. The standard approach is to posit a covert distributivity operator, essentially with the semantics of adverbial *each* (Lasersohn, 1995; Link, 1991; Roberts, 1987). On this view, predicates like *sang happy birthday* are ambiguous between two readings because they are ambiguous between two structures: one without the distributivity operator $\mathcal{D}$ and one with it, as in (130).

(130)   a. [$_{VP}$ sang happy birthday]

   b. [$_{VP}$ $\mathcal{D}$ [$_{VP}$ sang happy birthday]]

There are many ways of implementing $\mathcal{D}$. For our purposes, we can assume that it has the same semantics as *each* and the L-distributivity meaning postulate (though see LaTerza (2014) for an event-based distributivity operator that can help explain further distinctions). Namely, $\mathcal{D}$ in (131) adjoins to the VP and takes two arguments: the verb phrase $\mathbf{V}$ and a second argument $\mathbf{X}$ that will be supplied by the subject when $[\mathcal{D} \ [_{VP} ...]]$ composes with it.

(131) $[\![\mathcal{D}]\!] = \lambda\mathbf{V}.\lambda\mathbf{X}.\forall y(\mathbf{X}y)[\mathbf{V}y]$

Given the semantics of $\mathcal{D}$ in (131), the distributive version of *sang happy birthday* in (130b) is represented as in (132).

(132) $\lambda\mathbf{X}.\forall y(\mathbf{X}y)[\mathbf{sang\text{-}happy\text{-}birthday}(y)]$

The distributive VP in (132) then serves as the $\mathbf{E}$ argument of a quantifier in a sentence like *each/every student sang happy birthday.* In the *each* case, the presence of $\mathcal{D}$ is redundant. The sentence already has a distributive interpretation by virtue of the overt *each.* This is captured by the fact that the inclusion of $\mathcal{D}$ leads to redundancy in the first-order (133b): *each thing$_x$ that satisfies* **Student** *is such that for each thing$_y$ identical to it$_x$, it$_y$ satisfies* **Sang-happy-birthday**.[6]

(133)   a. Each student [sang happy birthday]

        $= \forall x(\mathbf{Student}(x))[\mathbf{Sang\text{-}happy\text{-}birthday}(x)]$

    b. Each student [$\mathcal{D}$ [sang happy birthday]]

        $= \forall x(\mathbf{Student}(x))[\forall y(y = x)[\mathbf{Sang\text{-}happy\text{-}birthday}(y)]$

---

[6] The $\mathbf{X}$ in (132) saturated by x. This yields $x(y)$, glossed *y is x.* For readability, "$y = x$" replaces "$x(y)$" in (133b). When a second-order variable saturates $\mathbf{X}$, yielding Xy, this would instead be glossed *y is (an) X* or *it$_y$ is one of them$_X$*, and this substitution would be inappropriate.

In either case, the predicate is forced to apply to each individual student. So whether or not the distributivity operator $\mathcal{D}$ is present, *each student sang happy birthday* only receives the distributive interpretation that there were as many events of singing happy birthday as there were students.

In the second-order case, the presence or absence of the distributive operator makes a difference. When $\mathcal{D}$ is absent, the collective reading is achieved, as in (134a): *the students$_G$ are such that they$_G$ satisfy* **Sang-happy-birthday**. But when $\mathcal{D}$ is present, the predicate is forced to apply to each student, as in (134b): *the students$_G$ are such that each one$_y$ of them$_G$ satisfies* **Sang-happy-birthday**.

(134)   a. Every student [sang happy birthday]

$$= \iota G(\forall x[Gx \equiv \mathbf{Student}(x)])[\mathbf{Sang\text{-}happy\text{-}birthday}(G)]$$

       b. Every student [$\mathcal{D}$ [sang happy birthday]]

$$= \iota G(\forall x[Gx \equiv \mathbf{Student}(x)])[\forall y(Gy)[\mathbf{Sang\text{-}happy\text{-}birthday}(y)]$$

In sum, the first-order/second-order distinction is compatible with the standard approach to P-distributivity (a covert distributive operator in the syntax) in addition to the standard approach to L-distributivity (distributive meaning postulates). To be sure, the above discussion has abstracted away from the meanings of VPs (including, importantly, event semantics) and is by no means a full account of universal quantifiers and distributivity. But importantly, *every* and *all* having second-order representations creates no new problems in this domain. Moreover, *each* having a first-order representation is enough to explain why it only gives rise to distributive interpretations.

3.2 The corresponding psychological distinction

Having established the logical distinction between first-order and second-order quantifiers, we can turn to the corresponding psychological distinction between a system of individual representation (object-files) and a system of group representation (ensembles). The rest of this chapter will argue that the logical and psychological distinctions correspond in the following sense: First-order quantification like "$\forall x(...x...)$" is an instruction to cognition to initiate object-file (i.e., individual) representations whereas second-order quantification like "$\iota G(...G...)$" is an instruction to cognition to initiate an ensemble (i.e., group) representation.

3.2.1 Object-file representations

An object-file is essentially a pointer to an individual object in the world. The term originated with Kahneman and Treisman (1984), which introduced the idea of a representation responsible for linking successive experiences of the same object over time. When watching a ball fly through the air, for example, we represent a single entity in motion, not merely a series of perceptually-similar time-slices.

To take an empirical example, Kahneman et al. (1992) asked participants to watch videos of boxes containing letters moving around the screen. Suppose box_1 started on the left and contained an "L" and box_2 started on the right and contained an "R". After participants took note of their contents, the boxes closed, occluding the letters. At this point, the boxes were perceptually identical. Then, box_1 and box_2 switched places and one of them opened, revealing the letter inside. Participants

were asked to recall whether that letter was one of the initial letters or a new one they hadn't seen before. Perhaps surprisingly, participants performed best when box_2, now on the left side, opened to reveal an "R". They performed worse if the letter appeared in its original spatial location but was contained within a different box, for example if box_2 opened to reveal an "L". This suggests that participants represented each box as an object, bound a feature to that object (e.g., "contains an R"), and held the object in working memory. That is, they represented each box as a separate object-file, despite the boxes being irrelevant to the task.

Since Kahneman et al. (1992), object-files have been well-researched both in adults and infants. Carey (2009) brings together these two literatures and argues that the system responsible for representing object-files is a system of core cognition.[7] As Carey discusses, this system exhibits three crucial properties.

The first such property is that object-files privilege spatio-temporal information over kind or property information. Intuitively, imagine someone looking up at the sky and exclaiming: *It's a bird; no, it's a plane; no, it's superman!* Their experience is not one of three separate objects but of a single object (*it*) that they view under three different concepts in rapid succession. Notice that this is still true if their reason for revising their claim is because they updated every aspect of their perceptual representation (e.g., because superman was flying closer to them).

_____

[7] Spelke (2000) identifies at least four systems of core cognition: systems for representing objects, actions, number, and space (see Spelke and Kinzler (2007) for a brief review). The contents of core cognition representations are often conceptual, in that they cannot be reduced to sensory primitives and they can play an inferential role in general reasoning. But the representations themselves are often created by perceptual input analyzers that are modular in the sense of Fodor (1983). Namely, these systems consist of highly structured domain-specific innate mechanisms. In this way, representations in core cognition straddle the boundary between perception and belief (Jenkin, 2020).

Empirically, the importance of spatio-temporal information for object-file representations can be seen clearly in multiple object tracking experiments (Pylyshyn, 2001; Pylyshyn and Storm, 1988). In this paradigm, participants are asked to track three or four objects moving around the screen while ignoring similar-looking distractor objects. Altering properties of the objects (e.g., changing their color, shape, or size) while they are in motion does not impact participants' tracking ability. Moreover, if one of the objects disappears, participants can accurately identify its last known position and direction of motion but not properties like its shape or color at the time of disappearance (Scholl et al., 1999). This suggests that object-files are created/indexed based on spatial location and that binding features to an existing object-file representation is a secondary process. Kahneman et al. (1992) describe this as a distinction between *seeing* and *identifying*; Xu and Chun (2009) similarly distinguish object *individuation* from object *identification* and specify different neural correlates of these processes.

The same privilege for spatio-temporal information can be seen in infants. For example, Xu and Carey (1996) showed 10-month-olds a scene in which one object – a red metal truck – emerged from behind an occluder on the left, lingered in view for a moment, and moved back behind the occluder. Then a second object – a blue rubber elephant – emerged from the occluder on the right, lingered in view for a moment, and retreated. Then the occluder was lifted. In one condition, a single object was revealed; in another condition, both objects were revealed. Surprisingly, infants didn't seem to care if there turned out to be only a single object (they showed the same looking-patterns in either condition as they did in baseline controls). This

suggests that they were unable to use the kind information (truck vs. elephant) or the property information (metal, red vs. rubber, blue) to infer that there were two objects behind the occluder. On the other hand, when infants were given a simple spatial cue – both the truck and the elephant were shown on-screen simultaneously before being hidden behind the occluder – they were able to infer the presence of two objects.

To be sure, while spatio-temporal information is privileged for individuating object-files, it is not the only way to trigger their representation. Xu (2002), for example, shows that 9-month-old infants succeed at the Xu and Carey (1996) task if the two objects are labeled upon emerging from behind the screen. Labeling even helped infants represent two object-files when the objects and labels were completely novel. At the same time, non-linguistic tones and emotional expressions paired with the two objects did not support individuation.

The second crucial property of object-files is that multiple object-files can be represented simultaneously, but only up to a working memory limit of three or four items. This is perhaps best exemplified by Feigenson and Carey (2005), which demonstrated that infants can represent up to three objects in parallel but catastrophically fail at four. In one of their tasks, 12-month-olds were shown some number of balls placed inside a box, one at a time. Infants were then allowed to search the box themselves, and they always retrieved one ball. Hidden from the infants' view, the experimenter surreptitiously removed all remaining balls from a trap door in the back of the box. Infants were then allowed to search the box again, with the dependent measure being the amount of time spent searching. As expected,

if infants initially saw only one ball placed in, they would not continue searching the box after retrieving it. If they initially saw two or three balls placed in, they would continue searching after retrieving the first ball. But if they initially saw four balls placed in, infants would not search the box for the remaining three. They treated a box containing four balls the same way they treated a box containing only one.

Feigenson and Carey (2005) further demonstrated infants' failure to simultaneously represent four individuals using a task introduced in Feigenson et al. (2002). The experimenter placed graham crackers into two opaque buckets while 11-month-olds watched from the other side of a small room. Infants were then allowed to crawl to one of the buckets to retrieve the crackers (the buckets were far enough apart that infants could only pick one). Participants in this study reliably chose the bucket with two crackers over the bucket with only one, and they reliably chose the bucket with three crackers over the bucket with two. But surprisingly, if given the choice between four crackers and one cracker, 11-month-olds were at chance, picking the bucket with only one cracker half of the time.

This result is especially impressive because the four-versus-one comparison is confounded with overall mass of cracker and with the time that the experimenter spent near each bucket. Infants fail to use these cues. Instead, they seem to be attempting to "open" a new object-file for each individual cracker, which puts them up against their working memory limit. Going beyond this limit seems to have catastrophic effects. We might have expected infants to open three object files and then stop (in which case they would succeed at a 4 vs. 1 comparison as if it were a 3 vs.1 comparison). But instead, infants seemingly attempt to open a fourth and end

up losing track of the initial three. Infants were able to succeed at this task when given the choice between four and zero crackers, suggesting that when their working memory capacity is exceeded, they nonetheless represent that there is something in the bucket. But they do not retain their original three object-files.

A similar working memory limit on representing object-files is present in adults. For example, a common result from the multiple object tracking paradigm mentioned above is that the task becomes very difficult when participants are required to track five or more objects. And results from change detection paradigms – in which participants are briefly shown two images of objects and asked to identify which object changed in the second image – show that performance sharply declines when there are more than four items in the initial display (Vogel et al., 2001). Of course, adults do not fail as catastrophically as infants. Presumably, this is because other strategies are available to us (e.g., verbally counting or refraining from even trying to create more object-files than we can reasonably track). So while there may not be a fixed upper limit of three items for adults, representing multiple object-files at once is capacity-limited in some sense (Alvarez and Franconeri, 2007).

The third crucial feature of the system for representing object-files is the capacity to track individual objects through occlusion. Scholl and Pylyshyn (1999) show that adult participants can successfully track multiple dots, even when they are briefly hidden by a barrier. In another condition, the dots instead "imploded" out of existence before subsequently "exploding" back into existence (i.e., the dot in question shrank until it disappeared completely; then, maintaining the same trajectory, a new dot grew from nothing). In this "implosion" case, tracking is interrupted

despite the fact that the dots to-be-tracked were removed from participants' visual fields for the same amount of time as in the occlusion case. Perhaps surprisingly, whether an object-file representation survives perceptual absence depends on how the object disappears.

Cheries et al. (2005) extended this finding to 10-month-olds. Infants who watched videos of two balls moving behind occluding barriers were surprised to see three balls when the barriers were removed in the test phase. Likewise, infants who watched videos of three balls moving behind occluding barriers were surprised to see two balls when the barriers were removed. But infants who watched balls in the "imploding" condition were not surprised to see a different number of balls in the test phase. In the same vein, Kaufman et al. (2005) show that 6-month-olds distinguish occlusion from disintegration and exhibit a brain response associated with representing objects only in the former case.

In sum, object-file representations are mental pointers to objects in the world. They individuate objects on the basis of spatio-temporal information and, after individuation, allow for binding of individual properties. They can be tracked through occlusion, but there is a limit to how many object-files can be represented at once, likely owing to working memory.

3.2.2   Ensemble representations

If an object-file is a pointer to an individual object, an ensemble can be thought of as a pointer to potentially many objects at once. Due to the capacity limits dis-

cussed above, indexing multiple objects at once requires abstracting away from the individuals and their particular properties. Instead, ensembles are represented in terms of summary statistics like cardinality, center of mass, average size, density, perimeter, average hue, average orientation, and so on (for helpful reviews, see: Alvarez 2011; Haberman and Whitney 2012; Whitney and Yamanashi Leib 2018). These summary statistics are rapidly extracted and do not require precise representation of the individuals over which they are computed.

To take an influential example, consider participants' ability to extract average size. Ariely (2001) showed participants displays of between four and sixteen different-sized circles for 500 milliseconds. After each display, they were immediately shown a single circle in the center of the screen. When asked to decide whether or not a circle with that particular size had been present in the previous display, participants performed essentially at chance. This effect was replicated with a two-alternative-forced-choice version, in which participants were at chance picking between two circles, one of which had been present in the previous display. But when asked whether a single circle was bigger or smaller than the mean circle size from the previous display, participants were able to offer surprisingly precise judgments.

Chong and Treisman (2003) provide further support for our impressive ability to extract the average size summary statistic. Even with exposure times as short as 50 milliseconds, participants in their experiments were able to judge which of two simultaneously presented arrays of 12 circles had a larger average size. They were able to do so just as well as they were able to judge which of two simultaneously presented circles was larger. A variety of control conditions ruled out alternative strategies

not based on computing the average (e.g., if a homogenous array of medium circles is pitted against a bimodal array of small and large circles, participants cannot arrive at the answer merely by matching individuals). Adapting this task for children, Sweeny et al. (2015) show that 4- to 5-year-olds can also rapidly extract average size. In particular, participants were able to decide which of two trees had larger oranges on average despite showing poor precision when asked to discriminate individual orange sizes.

It seems then, that even in the absence of representing individuals as such, participants are able to represent the entire group of circles in terms of the summary statistic average size. Based on comparisons of the variance of estimates of average size and estimates of individual size, Im and Halberda (2013) moreover argue that participants do not compute the summary statistic by sampling from the individuals constituting the ensemble. Instead, they suggest that extracting a summary statistic like average size is akin to texture processing.

In addition to point-estimates like average size, ensemble summary statistics also include measures of the range. For example, while Ariely (2001) found that participants were generally bad at discerning whether a particular circle was present in a display, they were able to correctly reject circles that were larger or smaller than any circle in the initial display. Demeyere et al. (2008) replicate this result in a patient with simultanagnosia, who is unable to focus his attention on individual objects (e.g., he cannot count more than one or two objects in complex displays, and he reports that he is only guessing despite showing signs of computing the average size). Likewise, Ward et al. (2016) showed that participants can accurately judge

the color diversity of an array of letters even in the absence of knowledge about the individual letters' colors.

Going beyond lower-level features, ensemble representations can be formed at higher-level of visual processing as well. A surprisingly counterintuitive case comes from Haberman and Whitney (2011), which presented participants with an array of sixteen greyscale faces with varying levels of happiness. After viewing this display for one second, participants were exposed to a second array of sixteen. Twelve of the faces were the same as in the initial display, but the four most emotionally extreme faces (i.e., the four happiest or the four saddest) were replaced with faces from the other end of the emotional spectrum. Participants were first asked to click on one of the four faces that changed. They were then asked whether the first or second array was happier on average. The most interesting cases are trials in which participants failed to correctly identify one of the four changes. On these trials, participants were nonetheless well above chance when asked about the average happiness. This result suggests that encoding individual properties is not a prerequisite for encoding ensemble summary statistics, even when the dimension in question is as high-level as emotion.

Two summary statistics in particular will be important for the experiments reported in section 3.3. The first is center of mass (which is also sometimes called centroid or mean location). In one study probing center of mass, Alvarez and Oliva (2008) asked participants to complete a multiple object tracking task with eight homogenous dots, four of which they were told to track. At a random point, the display was masked for 200 milliseconds before either the four target dots or the

four distractor dots returned. Participants' task was to identify the center of mass of the missing group of four. They performed well above chance in either case. Surprisingly, their performance was identical regardless of whether they were asked to recall the center of mass of the group being tracked or the distractor group. In contrast, when seven dots reappeared and participants were asked to identify the location of the missing eighth dot, they showed better performance if that dot was one of the targets. This again demonstrates that ensemble summary statistics can be represented even when individual features (of the sort that get bound to object-files) are not.

The second particularly relevant summary statistic is cardinality. Estimates of cardinality are supported by the Approximate Number System, which was discussed in Chapter 2 (section 2.2.2). Cardinality is unlike the other summary statistics discussed so far in that it is not an average or a measure of variability. As Alvarez (2011) points out though, cardinality, like other summary statistics, can be extracted rapidly from multiple groups simultaneously. In particular, Halberda et al. (2006) show that participants suffer no decrement in performance when three groups of dots (including the superset) are flashed on screen simultaneously. For example, participants shown red dots and blue dots can reliably report the cardinality of the red dots, the blue dots, and the superset of all dots just as well as if there had been only a single color.[8] Zosh et al. (2011) replicate this result in 9-month-old infants.

---

[8] At four groups, performance begins to deteriorate. This suggests that while the number of individual objects that go into one ensemble is unlimited, the number of ensembles that can be simultaneously represented is limited to three. Interestingly though, the superset is always one of the ensembles represented. That is, even when six colors were present, adult participants in Halberda et al. (2006) accurately estimated the number of all dots. And even when four colors were present, infants in Zosh et al. (2011) noticed a change in the total number of dots on screen.

Moreover, Burr and Ross (2008) argue that cardinality, like other summary statistics, is a "primary visual property." In particular, they show that number is adaptable: repeatedly presenting participants with large numerosities in one area of their visual field led to vastly increased underestimation of arrays presented in the area. These findings motivate including approximate cardinality among the ensemble summary statistics.[9]

One final caveat is in order. Unlike with object-files, there is not as strong of a consensus in the literature about how exactly to understand ensembles. At the start of this section, ensemble representations were described as "pointers to potentially many objects at once." This description makes the analogy to second-order variables as plural variables especially clear (as discussed in section 3.1, the plural logic way to understand "$\exists X(...X...)$" is *there exists one or more things$_X$ such that they$_X$...*). However, it might turn out that ensemble representations are better understood as analogous to representations of sets.[10] Given that nothing in the rest of this chapter turns on the distinction, we can remain agnostic on this question.

That said, there are reasons for doubting the idea that representing an ensemble should be understood as representing a set. For one thing, sets preserve the identity of their members whereas the hallmark of ensembles is that they abstract

---

[9] An alternative to approximate cardinality being an ensemble summary statistic is that the ANS takes ensemble representations as input and uses them to generate cardinality estimates. This is an interesting distinction, but it is not one that has any bearing on the experiments in section 3.3.1. There, cardinality knowledge is used as a signal of ensemble representation. Whether cardinality is an ensemble summary statistic or ensembles are a prerequisite for obtaining cardinality estimates, the link between the two seems clear.

[10] Some authors colloquially refer to ensembles as sets, including Ariely (2001) and Halberda et al. (2006). However, nowhere in the literature on ensembles is an argument made for understanding ensembles in terms of the mathematical notion of a set.

away from individual-properties.[11]

Moreover, it is not clearly sensible to talk about a set as having a center of mass or a perimeter (as opposed to saying that its members have a center of mass or perimeter). The same is true for many other summary statistics. Cardinality is one exception: It makes sense to talk about a set having a cardinality, but only because the cardinality of any set is one. The relevant notion of cardinality – the one that plays the role of a summary statistic in an ensemble representation – is the cardinality of the set's members, not that of the set itself. But if all summary statistics are computed over the members, is there any theoretical role left for the set itself to play?

3.3   Psychosemantic evidence: How are arguments represented?

In Chapter 2, we used cardinality-knowledge as a proxy for argument representation and saw that evaluating sentences with *every* or *all* leads participants to represent the extension of the determiner's first argument. But cardinality-knowledge signals more than just representing an argument. It also signals that the argument in question was represented in a particular way, namely, as an ensemble as opposed to as independent object-files. Here, we will see that *every* leads to better cardinality-knowledge (section 3.3.1) and center-of-mass-knowledge (section 3.3.3) than *each*, but that *each* leads to better individual-color-knowledge than *every* (section 3.3.2). This suggests that while participants treat the first argument of *every* and *all* as an

---

[11] They do so likely because of working memory constraints: Representing a plurality by representing each individual would be too costly.

ensemble, they treat the first argument of *each* as independent object-files.

There are likely many factors that go into determining whether someone will treat the extension of a quantifier's first argument as an ensemble or a number of independent object-files. Low-level visual properties of the scene will matter (e.g., spreading out objects to an extreme degree will probably discourage ensemble representation), as will higher-level properties (e.g., grouping five lemons into an ensemble is more natural than grouping two lemons, a fruit bowl, a cutting board, and a knife).

A central claim of this chapter is that another factor that goes into determining how an argument is represented is whether the meaning representation is first-order or second-order. In particular, the claim here is that a representation that contains a first-order quantifier like "$\forall x(...x...)$" provides an instruction to create individual object-file representations (in this case, of each x). On the other hand, a representation that contains a second-order quantifier like "$\iota G(...G...)$" provides an instruction to create an ensemble representation (in this case, of the Gs). These are two instantiations of the Interface Transparency Thesis (ITT) in (135), as discussed in Chapter 1 (section 1.3.1).

(135) Linking hypothesis (ITT): The verification procedures employed in understanding a declarative sentence are biased towards algorithms that directly compute the relations and operations expressed by the semantic representation of that sentence (Lidz et al., 2011).

     a. first-order semantic representation $\rightarrow$ object-files

b. second-order semantic representation → ensemble

Given these linking hypotheses, the results of the experiments reported below support the pair of lexical semantic hypotheses in (136).

(136) Lexical semantic hypotheses

a. *each* has a first-order representation: $\lambda\mathbf{I}.\lambda\mathbf{E}.\forall x(\mathbf{I}x)[\mathbf{E}x]$

b. *every* has a second-order representation: $\lambda\mathbf{I}.\lambda\mathbf{E}.\iota G(\forall x[Gx \equiv \mathbf{I}x])[\mathbf{E}(G)]$

The methodological strategy in these experiments will be to hold the scenes and truth-conditions (and sometimes the participants) constant and vary only the determiner. Otherwise unexplained variation in verification can then reasonably be attributed to the semantic representations under evaluation. In particular, (136a) predicts that, all else equal, when evaluating a sentence like *each circle is green*, participants will treat the circles as independent object-files. Conversely, (136b) predicts that, all else equal, when evaluating a sentence like *every circle is green* participants will treat the circles as a single ensemble.

Spelling out the exact details of the object-file-based and ensemble-based verification strategies that participants use to evaluate universally quantified statements is a topic for future work. In the object-file case, it might involve cycling through each individual object. In the ensemble-case, one possibility is that participants rely on summary statistics indexing the level of homogeneity, like color diversity (Ward et al., 2016). For example, if *every big circle is blue* is TRUE, then the big circles will have a low level of color diversity. But if *every big circle is blue* is FALSE, then

the big circles will include a larger range of hues.[12] For present purposes, we can remain agnostic about the precise details of the strategy and focus on evidence that participants represent ensembles or object-files.

3.3.1   Experiments 1 & 2: Probing cardinality knowledge in adults

These experiments were first reported in Knowlton et al. (accepted). Similar to the experiments in Chapter 2, participants here were shown dot-displays like the one in Figure 3.1. They were asked to evaluate sentences like *every big dot is blue.* After answering TRUE or FALSE, the image disappeared, and they were asked to recall the cardinality of some group of dots from the display. Sometimes they were asked about a distractor set, which was not mentioned in the initial sentence (small or medium dots, in this example). Other times they were asked about the target set (i.e., the one denoted by the quantifier's internal argument; big dots, in this case).

Regardless of details about the quantifier's representation, performance on cardinality questions is predicted to be better when those questions probe the target set than when they probe a distractor set (if only because the former was mentioned and is relevant for arriving at the answer). However, a first-order representation, like the one proposed for *each*, should, by hypothesis, lead to less ensemble representation compared to a second-order representation, like the one proposed for

---

[12] Given that some color categories contain more hues than others, this strategy raises an interesting question: Will participants be more likely to tolerate exceptions for smaller color categories? Or are they able to adjust the acceptable range depending on the predicate? It is also a possibility that participants use the range as a first-pass measure. With a small range and and an average solidly within the desired category, the sentence is likely TRUE; but a large range may necessitate looking for disconfirming instances. This might predict, for example, longer reaction times on FALSE trials with large ranges. Thanks to Sandy LaTourrette for helpful discussion on this point.

every. Consequently, evaluating a sentence with *each* is predicted to lead to worse performance when asked to recall the cardinality of the target set, compared to evaluating an otherwise identical sentence with *every* or *all.* Experiment 1 finds this advantage for *every* over *each* (section 3.3.1.1). In a separate group of participants, Experiment 2 finds similar (though less conclusive) results when *every* is replaced with *all* (section 3.3.1.2).
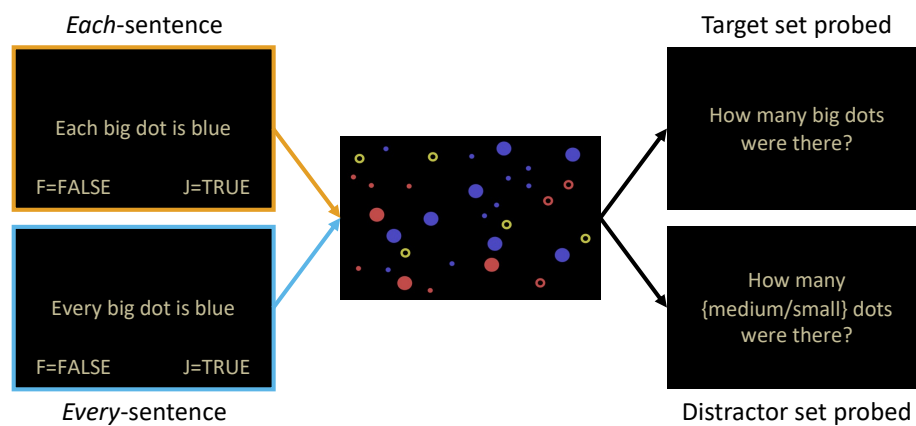
Figure 3.1: Trial structure of Experiment 1. Both manipulations – quantifier (*each* vs. *every*) and set probed (target vs. distractor) – were within-subjects, but quantifier was blocked.

3.3.1.1 Experiment 1: Cardinality, *Each* vs. *Every*

Participants: 30 University of Maryland undergraduates participated for course credit. All were native speakers of English. Participants were removed if they (a) scored below 85% accuracy on the TRUE/FALSE portion of the task, (b) reported that they used an explicit strategy, or (c) failed to complete both blocks of the experiment in the allotted hour. In this case, two participants were excluded from analysis for reason (b) and four were excluded for reason (c), leaving 24 participants.

Materials: Sentences had one of two forms, given in (137).

(137)   a. Each {big/medium/small} dot is {red/yellow/blue}

       b. Every {big/medium/small} dot is {red/yellow/blue}

Half of these sentences were TRUE with respect to the display and half were FALSE. Dot-displays consisted of a black background with red, yellow, and blue dots that could be big, medium, or small (e.g., Figure 3.1). Medium dots had black holes in the middle, to make them more distinguishable from the other two sizes (Chen, 1982, 2005). The dot sizes and colors were named during the training portion of the experiment to ensure that they could be correctly identified by participants (and participants were happy to refer to the circles with holes as "dots").

Each display contained between 24 and 48 dots. Out of the nine possible size/color combinations (e.g., big blue dots), six were always present. Each size/color combination present contained a minimum of 3 and a maximum of 9 dots. One difference between these experiments and those in Chapter 2 is that here, no follow-up questions probed the conjunction of both predicates. Instead, they were distributed as follows: 30 questions probed the target size, 30 probed a distractor size, 30 probed the target color, 30 probed a distractor color, and 16 probed the total number of dots. This led to 136 trials for each quantifier (272 total trials per participant). Only the target and distractor size questions were analyzed (the others were included as fillers).

Procedure: The experiment was conducted in a small, sound-attenuated room at the University of Maryland. On each trial, participants read a quantifica-

tional sentence. Quantifier was blocked and the initial condition was counterbalanced, such that half of the participants started with *each*-sentences and half of the participants started with *every*-sentences. After reading the sentence and pressing "space," they viewed a dot-display and evaluated whether the sentence was TRUE or FALSE with respect to the display by pressing "J" or "F" on their keyboard. Displays remained on-screen until participants offered their judgement (though they were told to respond as quickly as possible, to discourage explicit counting). After offering their TRUE/FALSE judgement, the screen went blank and participants were asked to recall the cardinality of one of the groups present in the display by typing in a number and pressing "enter". Participants were allowed a short break between blocks.

Predictions: As discussed above, distractor questions are predicted to lead to worse overall performance than target questions. If asked whether *every big dot is blue*, for example, participants have no reason to know the number of small dots. These distractor questions serve as a baseline measure of poorest possible cardinality estimation performance.

If *every* has a second-order representation, participants should be biased to represent the big dots as an ensemble. As a consequence, they should perform well on target questions. If *each* has a first-order representation, participants should be biased to represent the big dots as object-files when evaluating a sentence like *each big dot is blue*. This bias to represent object-files should result in worse performance on target questions than *every*. So despite the fact that these sentences are truth-conditionally equivalent, participants should, by hypothesis, have better estimates

of the cardinality of the internal argument (e.g., big dots) after evaluating the *every*-variant than after evaluating the *each*-variant.

Results: On the TRUE/FALSE portion of the task, participants correctly evaluated 96.9% of the *each*-sentences and 97.2% of the *every*-sentences. Time spent viewing the display in the two blocks did not significantly differ ($t = 0.55, p = .589$).

Participants' responses on the cardinality portion of the task bore out the predictions above. Figure 3.2 plots participants' performance in terms of percent error (lower error indicates better performance).
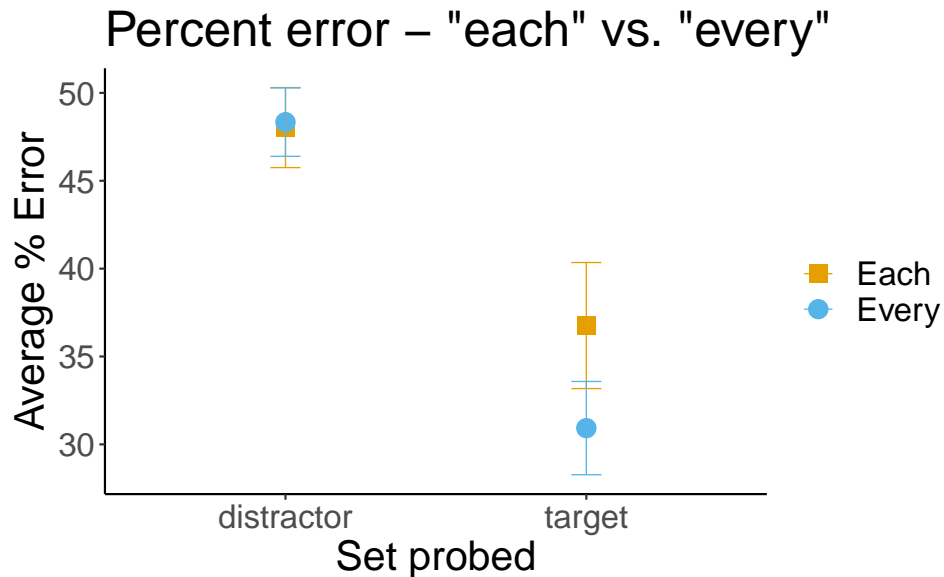
Figure 3.2: Percent error on cardinality questions in Experiment 1. Higher percent error reflects poorer cardinality estimations. For example, if the actual number of dots shown was 10, a response of 8 or 12 would result in 20% error; a response of 6 or 14 would result in 40% error.

To analyze responses on the cardinality portion, data were fit to different versions of the standard cardinality estimation model, discussed in Chapter 2 (see also Odic et al. (2016) and citations therein). To reiterate: Numerical estimates (y)

in response to being shown some number of objects (x) are modeled as a Gaussian distribution. The mean of this distribution is the number of objects (x) raised to the power of participants' accuracy (β) and multiplied by a scaling factor (α). The standard deviation scales linearly with the mean at a rate given by participants' precision (σ). This yields the basic model in (138).

(138) Basic cardinality estimation model:

$$y \sim \mathcal{N} \left(\begin{array}{c} mean = \alpha x^{\beta} \\ stdev = \sigma \times mean \end{array} \right)$$

Building off of (138), the models in (139) allow the scaling factor, accuracy, and precision to vary as a function of either the set probed, the quantifier, or both.[13] The "set probed only" model, (139a), allows the values of α, β, and σ to vary only based on the set probed (for target: $S = 1$; for distractor $S = 0$). The "main effects" model, (139b), allows these values to vary based on the set probed and the quantifier (for *every* $Q = 1$; for *each* $Q = 0$). The "interaction" model, (139c), allows them to vary based on the set probed and an interaction between the set probed and the quantifier. Lastly, the "full" model, (139d), includes terms for a main effect of set probed, a main effect of quantifier, and their interaction.

(139)   a. Set probed only:

$$y \sim \mathcal{N} \left(\begin{array}{c} mean = (\alpha_0 + \alpha_1 S)x^{\beta_0 + \beta_1 S} \\ stdev = (\sigma_0 + \sigma_1 S) \times mean \end{array} \right)$$

---

[13] This is a different analytical strategy than the one used in Chapter 2, where accuracies were independently fit and compared against a baseline. The absence of an ideal-performance baseline for these experiments makes that approach impossible. However, the same pattern of results is obtained if accuracies from the trials in each cell of the 2×2 design are computed and compared: the β value of a model fit to *every* target trials is higher than the β value for *each* target trials, and both are higher than the β values for distractor trials.

b. Main effects:

$$y \sim \mathcal{N} \left(\begin{array}{l} mean = (\alpha_0 + \alpha_1 S + \alpha_2 Q) x^{\beta_0 + \beta_1 S + \beta_2 Q} \\ stdev = (\sigma_0 + \sigma_1 S + \sigma_2 Q) \times mean \end{array} \right)$$

c. Interaction:

$$y \sim \mathcal{N} \left(\begin{array}{l} mean = (\alpha_0 + \alpha_1 S + \alpha_3 SQ) x^{\beta_0 + \beta_1 S + \beta_3 SQ} \\ stdev = (\sigma_0 + \sigma_1 S + \sigma_3 SQ) \times mean \end{array} \right)$$

d. Full:

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

The set probed only model in (139a) corresponds to the null hypothesis that the quantifier used has no effect on participants' cardinality estimates. The other three models are three different ways of spelling out the alternative hypothesis that the quantifier does have an impact on participants' cardinality estimation ability. The "interaction" model in (139c) is most in line with the predictions outlined above. Namely, participants should perform better on target questions than on distractor questions regardless of quantifier, but *every*-sentences should lead to better performance on target questions than *each*-sentences.

Models were fit using maximum likelihood estimation. To compare them, we can consider their relative Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values, both of which reward models for capturing the data and penalize them for including greater numbers of parameters (Akaike, 1998; Schwarz, 1978; Stone, 1979). Lower values are indicative of striking a better trade-off between fit and complexity. As seen in Table 3.1, both the AIC and BIC values

favor the predicated interaction model. This means that regardless of quantifier, participants had better knowledge of cardinality of groups that were mentioned in the sentence. But more importantly, *every*-sentences lead to better estimates of the internal argument's cardinality than truth-conditionally equivalent *each*-sentences.

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

Table 3.1: Model comparisons for Experiment 1. Lowest values are in bold.

The result is most pronounced among the 12 participants who completed the *each* condition first, as seen in Figure 3.3 (Table 3.2 likewise shows a same statistical result in the same direction as the whole-group analysis).

Given that *every*, in this task, leads to a superior strategy – in that it leads to better responses to subsequent cardinality questions – it would come as no surprise if participants who started in that condition were sometimes prone to retain this strategy upon encountering *each* instead of switching to an inferior alternative. Considering the *each*-initial participants, though, allows us to get a better sense of participants' unadulterated performance. That is, these participants were not linguistically biased to use an ensemble-based strategy prior to completing the *each* block (in which sub-optimal performance was predicted).

Overall, these results are well-explained if *every* (but not *each*) has a second-order representation and if second-order representations in turn encourage ensemble representation.

Figure 3.3: Percent error on cardinality questions for Experiment 1 participants who completed the *each* condition first. Higher percent error reflects poorer cardinality estimations. For example, if the actual number of dots shown was 10, a response of 8 or 12 would result in 20% error; a response of 6 or 14 would result in 40% error.

Model	AIC	BIC
Set probed only	8001.43	8033.00
Main effects	7918.65	7966.00
Interaction	**7873.30**	**7920.65**
Full	7879.27	7942.41

Table 3.2: Model comparisons for Experiment 1 participants who completed the *each* condition first. Lowest values are in bold.

3.3.1.2 Experiment 2: Cardinality, *Each* vs. *All*

Participants: 27 University of Maryland undergraduates participated for course credit. All were native speakers of English. Participants were removed following the same criteria as in Experiment 1. In this case, three participants were excluded for being unable to complete the experiment in the allotted hour, leaving 24 participants.

Materials: Materials were identical to Experiment 1 save for the change from *every*-sentences to the *all*-sentences in (140b).

(140)   a. Each {big/medium/small} dot is {red/yellow/blue}

   b. All {big/medium/small} dots are {red/yellow/blue}

It is worth noting that this change introduces a confound that was not present in Experiment 1: plurality. Because the plural *dots* may also be reasonably expected to encourage representing the relevant dots as an ensemble, the sentences in (140) do not constitute a true minimal pair.

Procedure: The procedure was identical to that of Experiment 1.

Predictions: As before, distractor questions are predicted to lead to worse overall performance than target questions. If *all*, like *every*, has a second-order representation, *all*-sentences should lead to better performance on target questions than *each*-sentences.

Results: On the TRUE/FALSE portion of the task, participants correctly evaluated 96.7% of the *all*-sentences and 96.4% of the *each*-sentences. Time spent viewing the display in the two blocks did not significantly differ ($t = 1.33, p = .197$).

Participants' responses on the cardinality portion of the task were analyzed in the same way as in Experiment 1. Figure 3.4 plots participants' performance in terms of percent error and Table 3.3 presents the model comparison values. Both the AIC and the BIC values favor the full model.

This suggests again that the difference in quantifier matters for cardinality estimation ability in the predicted direction of *all* leading to better performance
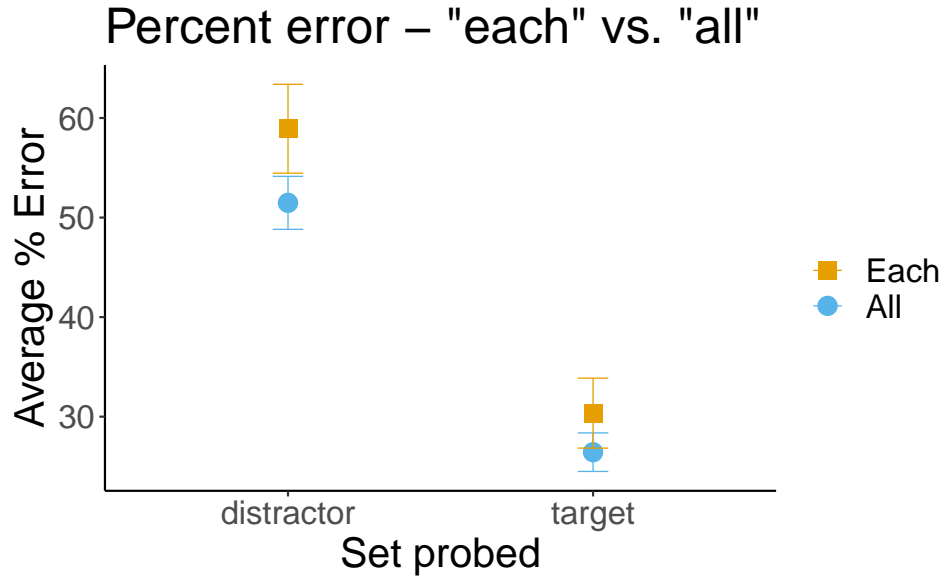
Figure 3.4: Percent error on cardinality questions in Experiment 2. Higher percent error reflects poorer cardinality estimations. For example, if the actual number of dots shown was 10, a response of 8 or 12 would result in 20% error; a response of 6 or 14 would result in 40% error.

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

Table 3.3: Model comparisons for Experiment 2. Lowest values are in bold.

than *each*. As in Experiment 1, this is well-explained if *all* (but not *each*) has a second-order representation and if second-order representations encourage ensemble representation. However, the observation that the full model is a better fit than the interaction model is surprising. This means that the benefit afforded by *all* is present even on distractor questions, where the set probed was not even mentioned in the initial sentence. However, this is likely a spurious result.

As with Experiment 1, the result is especially pronounced among participants

who completed the *each* condition first, as seen in Figure 3.5. Moreover, as seen in Table 3.4, while one measure of model comparison prefers the full model (as in the full sample), the other prefers the interaction model (as in Experiment 1).
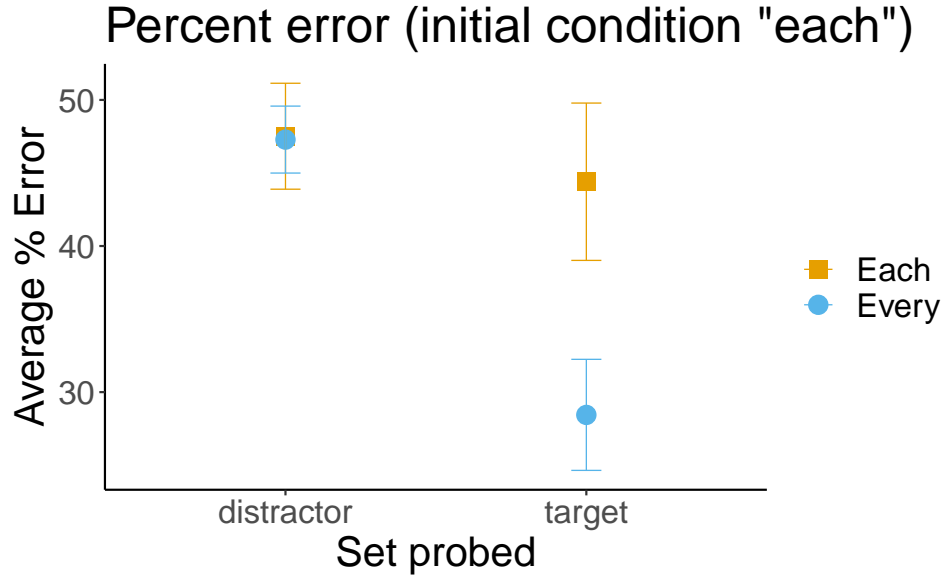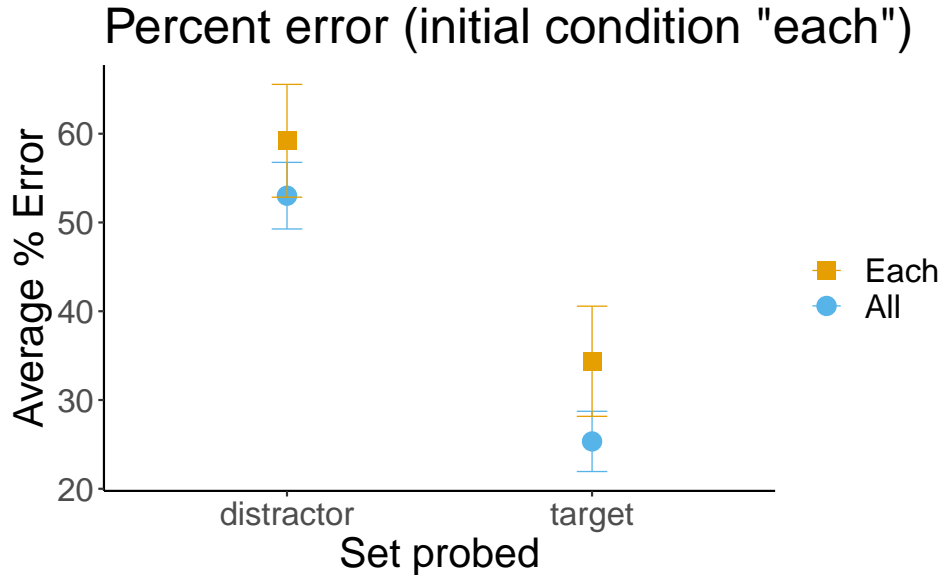

Figure 3.5: Percent error on cardinality questions for Experiment 2 participants who completed the *each* condition first. Higher percent error reflects poorer cardinality estimations. For example, if the actual number of dots shown was 10, a response of 8 or 12 would result in 20% error; a response of 6 or 14 would result in 40% error.

Model	AIC	BIC
Set probed only	7997.06	8028.56
Main effects	7950.58	7997.82
Interaction	7945.04	**7992.28**
Full	**7941.99**	8004.98

Table 3.4: Model comparisons for Experiment 2 participants who completed the *each* condition first. Lowest values are in bold.

Broadly speaking, the results of these first two experiments suggest that *every*-sentences (and *all*-sentences) encourage participants to represent the determiner's internal argument as an ensemble more often than truth-conditionally equivalent

each-sentences do. Given the linking hypothesis discussed above, these findings support the lexical semantic hypotheses outlined above: Both *every* and *all* have second-order representations whereas *each* has a first-order representation.

3.3.2 Experiments 3 & 4: Probing color knowledge in adults

The previous two experiments probed cardinality, an ensemble summary statistic. Experiments 3 and 4 instead probe a property that can be bound to object-files, individual color. Participants were shown displays containing three circles, each with a different hue, as in Figure 3.6. They were asked to evaluate sentences like *each/every circle is green.* After answering TRUE or FALSE, the circles disappeared briefly before reappearing with one circle having potentially changed its hue. Participants' task was to say whether they detected the change.
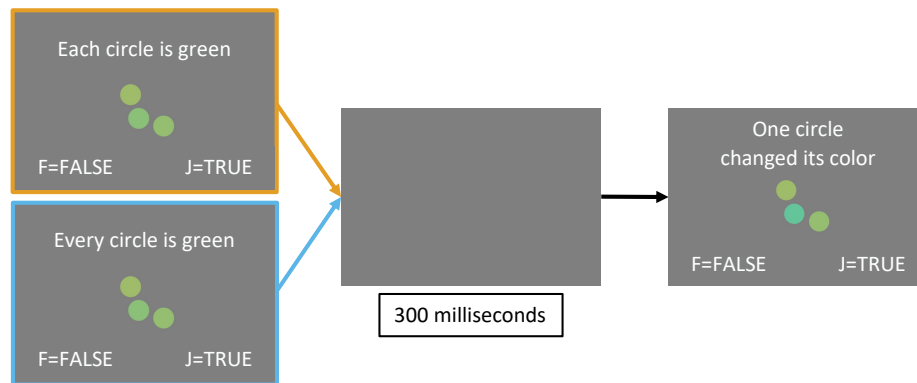
Figure 3.6: Trial structure of Experiment 3. Participants were always shown three circles, each of which was a unique shade of green, orange, or blue. In this case the follow-up statement is TRUE because the middle circle is a different hue. Quantifier was varied between-subjects.

Given that ensembles abstract away from individuals and their properties,

representing the circles in this task as an ensemble should lead to inferior performance on the subsequent change-detection task relative to representing the circles as object-files. So while the experiment in section 3.3.1.1 found superior performance on an ensemble-based follow-up question for *every*, the experiments here predict superior performance on an individual-based follow-up question for *each*. Experiment 3 finds this difference in a standard change detection task (section 3.3.2.1). Experiment 4 replicates this finding in a staircased change detection task, where the difficulty varies as a function of participants' performance (section 3.3.2.2)

3.3.2.1 Experiment 3: Color, constant difficulty

Participants: 43 participants were recruited online using Amazon Mechanical Turk. All passed an English-screener prior to the actual task. Seven participants were removed from further analysis either for achieving below 50% accuracy on the TRUE/FALSE portion of the task (five) or for having response times longer than three standard deviations above the mean response time (two). This left 36 participants.

Materials: Sentences had one of two forms, given in (141).

(141)   a. Each circle is {green/blue/orange}

b. Every circle is {green/blue/orange}

Unlike Experiments 1 and 2, quantifier was a between-subjects manipulation in Experiments 2 and 3. Half of the participants evaluated *each*-sentences and the other half evaluated *every*-sentences.

Displays consisted of a grey background with three circles. The circles' colors

were randomly selected from an independently normed color wheel with a constant luminance and 180 equally-spaced hues (Bae et al., 2015). According to empirically-determined color categories, in half of the trials the sentences were TRUE with respect to the display and in half of the trials they were FALSE.[14]

For the follow-up change detection task, half of the trials were "no change" trials, in which the original three circles reappeared after a 300 millisecond grey screen. Half of the trials were "change" trials, in which one circle changed its hue in second display (after the 300ms grey screen). In these cases, one circle was randomly selected to change. A new color was then sampled from a Gaussian distribution with a mean of the original hue and a standard deviation of 17 (selection of the original hue was not permitted, however). Participants completed 84 trials in total.

Procedure: During the instructions, participants were led to believe the task was about color categories. On each trial, they read a quantificational sentence presented alongside three circles. They were given as long as they wanted to inspect the circles and decide whether they agreed with the sentence. After they evaluated the sentence as TRUE or FALSE by pressing "J" or "F" on their keyboard, the display disappeared for 300 milliseconds. Then three circles reappeared in the same spatial positions, alongside the text *One circle changed its color*, which participants could judge as TRUE or FALSE by pressing "J" or "F". Participants had unlimited time to respond.

---

[14] To be sure, participants differ in their judgments about color category boundaries. Some hues in Bae et al. (2015) were overwhelmingly categorized as a particular color (e.g., hue 83 was called *green* by all 60 participants asked) whereas others were borderline (e.g., hue 93 was judged to be green by 30 participants and to be blue by the other 30 participants). For these experiments, we considered a hue a member of a given color category if it was the modal response when adult participants were asked to name it.

It is potentially worth noting that the wording of this follow-up statement leaves the question somewhat open. In particular, *color* could be understood to mean *color category*, in which case many deviations in hue will still elicit the response FALSE, or *particular hue*, in which case any deviation at all makes the statement TRUE. Because colors were randomly sampled, some trials had changes that crossed color category boundaries whereas others had changes taken from within the same category. This did not seem to confuse participants, as they were generally willing to answer TRUE to within-category changes. That said, in subsequent versions of this task (not included in this dissertation) participants were warned that sometimes the changes would be large and other times the changes would be small.

Predictions: As discussed above, if *each* has a first-order representation (and if first-order representations are instructions to represent object-files), participants should be biased to treat the circles as individual object-files. Likewise, if *every* has a second-order representation (and if second-order representations are instructions to represent ensembles), participants should be biased to abstract away from the individual circles and treat them as an ensemble. As a consequence, participants should show better performance on the change detection task after evaluating an *each*-sentence than a truth-conditionally equivalent *every*-sentence.

Results: On the TRUE/FALSE portion of the task, participants correctly evaluated 67% of the *each*-sentences and 69.7% of the *every*-sentences (these values appear low, though see note 14). Time spent viewing the display did not significantly differ across conditions ($t = 0.71, p = .486$).

Participants' accuracy in the change detection portion of the task is plotted in
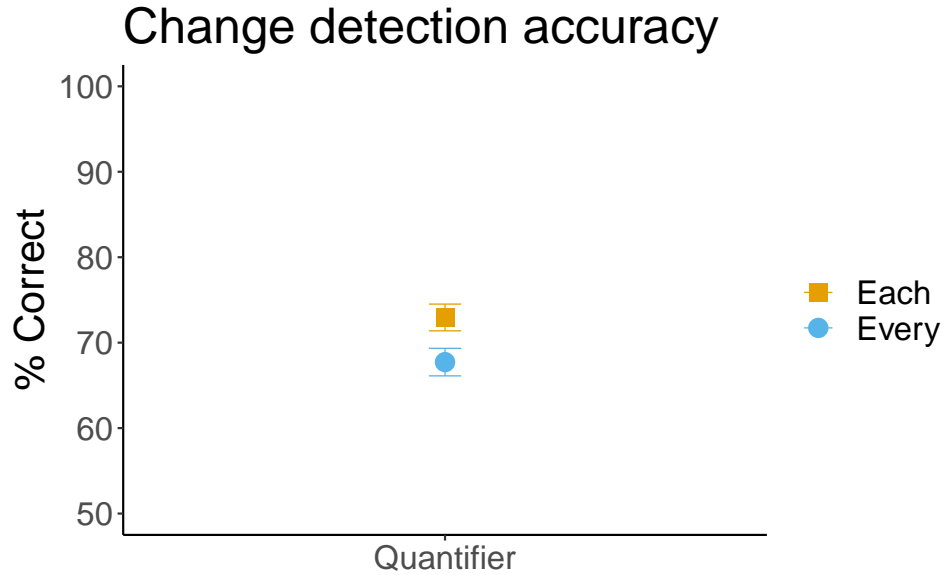
Figure 3.7: Percent correct on the change detection question in Experiment 3.

3.7. In line with the predictions discussed above, participants who first evaluated an *each*-sentence were more accurate than participants who first evaluated an *every*-sentence $(t = 2.34, p < .05)$.

3.3.2.2 Experiment 4: Color, staircased difficulty

Participants: 37 participants were recruited online using Amazon Mechanical Turk. All were required to pass an English-screener before participating in the actual task. One participant was removed from further analysis for having response times longer than three standard deviations above the mean response time. This left 36 participants.

Materials: Materials were identical to Experiment 3 with one exception: the distribution from which new colors were chosen differed. Instead of relying on a constant standard deviation (17 in Experiment 3), the standard deviation of this

distribution started at 20. 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). 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).

Procedure: The procedure was identical to that of Experiment 3.

Results: On the TRUE/FALSE portion of the task, participants correctly evaluated 65.9% of the *each*-sentences and 65.4% of the *every*-sentences. Time spent viewing the display did not significantly differ across conditions ($t = 1.73, p = .093$).

Because the difficulty of the change detection task varied based on performance, participants in both conditions achieved the same accuracy (*each*: 71.8%; *every*: 71.2%). Instead, the dependent measure of interest is the difficulty level required to achieve this degree of accuracy. As seen in Figure 3.8, participants who evaluated *each*-sentences on average had a smaller standard deviation – reflecting harder trials – than participants who evaluated *every*-sentences ($t = 11.65, p < .001$). That is, for participants to achieve 70% accuracy, change detection trials following *every* had to be easier than those following *each*.

In sum, Experiments 3 and 4 tested the prediction that *each*-sentences will lead to object-file representations, and consequently, memory for individual properties like color, whereas *every*-sentences will lead to ensemble representation, and consequently, degraded memory for individual properties. These predictions were borne out. In line with Experiments 1 and 2, these results are well-explained if *each* has a first-order representation and *every* has a second-order representation.

SD of new color distribution
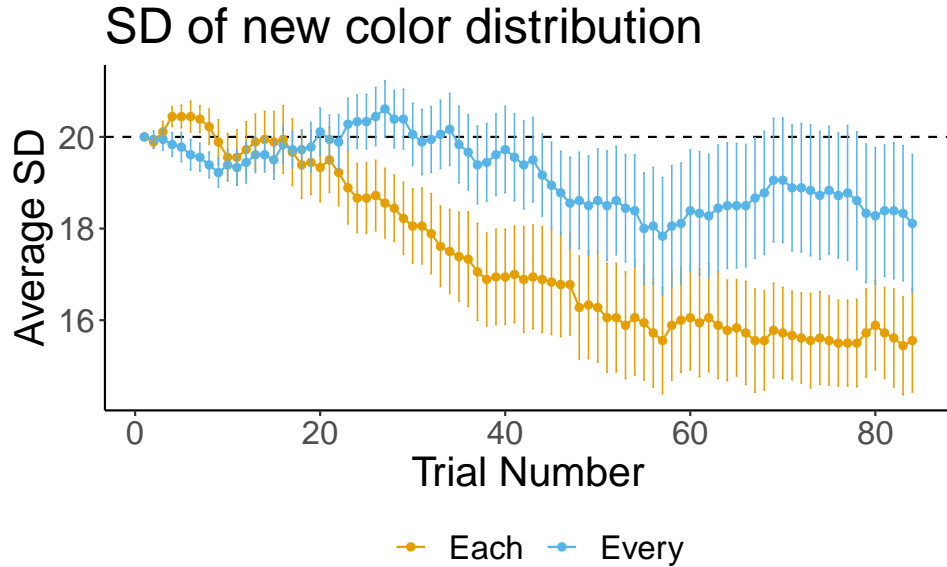
Figure 3.8: The standard deviation of the distribution from which the new color was chosen on each trial. A larger standard deviation corresponds to an easier trial, on average (since the new hue is less likely to be similar to the original). All participants started the task with the SD set to 20.

3.3.3 Experiment 5: Probing center of mass knowledge in children

Experiments 1-4 show that adults have a bias to use different verification strategies to evaluate *each*- and *every*-sentences, even when the scene remains constant. This variation in verification strategies was taken to reflect a representational difference. That difference amounts to the claim that knowing the meaning of *every* consists in having paired its pronunciation with a second-order universal concept and knowing the meaning of *each* consists in having paired its pronunciation with a first-order universal concept. A strong test of this conclusion would be to show that the same strategies used by adults are also deployed by children as early as they can be said to know the meanings of *each* and *every*.

Experiment 5 tests this prediction with a task appropriate for a wide age

range. Participants (3- to 8-years-old) were shown an iPad with a display like the one in Figure 3.9. They were either asked whether *each* or *every* circle was blue. After answering "yes" or "no", the circles disappeared and participants were asked to remember where the center of the circles was. They could record their guess by tapping a location on the iPad.

"Is each circle blue?"

"Is every circle blue?"

"Where was the middle of the circles?"
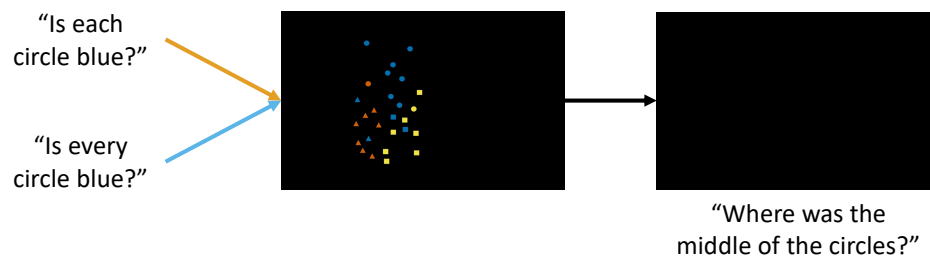
Figure 3.9: Trial structure of Experiment 5. Quantifier was varied between subjects. The center of the circles was always offset from the center of the screen in both the x and y directions.

Since center of mass is an ensemble summary statistic, children who answered the *every*-question should have better estimates of the center of mass than children who answered the *each*-question.

Participants: 175 children between 3- and 8-years-old participated. They were recruited at the University of Maryland and Johns Hopkins University. Prior to any analysis, 25 children were excluded for the following reasons: needing to tap the screen multiple times to record their guess (14), parental/sibling interference (5), failing to understand the task (5), or equipment failure (1). This left 150 children. Of these, 41 incorrectly answered the initial question and were removed from further

analysis (as they either did not yet know the meaning of the quantifier or were not paying attention to the task). This left 109 children (age range: 3;2 - 7;11, mean age: 5;8).

Materials: Each child participated in a single trial. Questions were identical save for the quantifier, which was manipulated between-subjects (see the full script in (142a)). Displays consisted of nine circles, nine triangles, and nine squares on a black background. There were always seven blue circles, one yellow circle, and one red circle. This means that the correct answer to the question *is each/every circle blue?* was always "no" (a correct rejection is a stronger test of knowledge, given children's "yes"-bias). Shapes appeared in a large cluster either on the left or right side of the screen to ensure that average position of the shapes and the average position of the circles was never the center of the screen. Eleven different images were randomly generated with these constraints and participants were shown one at random (meaning the *each* and *every* conditions used identical images).

Procedure: Participants were shown an all-black screen on an iPad. The researcher then followed the script in (142a). After the participant answered "yes" or "no", the researcher agreed with their response before advancing to the blank screen and following the script in (142b).

(142)   a. Let's play a game! I'm going to show you a picture with different shapes. Can you say whether {each/every} circle is blue? Are you ready?

<div align="center">CLICK TO REVEAL SHAPES</div>

Is {each/every} circle blue? What do you think? Is {each/every} circle blue? (repeat again if child does not respond.)

<div align="center">AGREE WITH RESPONSE AND CLICK TO REMOVE SHAPES</div>

b. Where was the middle of the circles? Can you tap where the middle of the circles was?

Predictions: As discussed above, if the different verification procedures for *each-* and *every*-sentences reflect the semantic representation of those sentences, then the asymmetry should be present as soon as the meaning of these expressions is known (i.e., as soon as learners have paired the pronunciation of *each* with a first-order universal concept and the pronunciation of *every* with a second-order universal concept). Because center of mass is an ensemble summary statistic, children in the *every* condition should perform better than those in the *each* condition.

On the other hand, children might go through a stage of, for example, thinking that *each* and *every* are synonymous and have both pronunciations paired with the same second-order concept. That is, they might understand that these determiners both express universal content before understanding exactly how that content is represented by adults. In this case, we should expect the effect of *every* leading to better performance than *each* not to be present at the youngest end of the age range, but to emerge over time (i.e., we expect an interaction between quantifier and age).

Results: As noted above, not all of the participants correctly answered the initial "yes"/"no" question (i.e., some incorrectly answered "yes"). Accuracy is broken down by age in Figure 3.10. The low accuracy at the lower end of our age range may result from some of the 3- to 5-year-olds not yet knowing the meaning of *each* or *every*. To be sure, the age of acquisition is still debated (see Chapter 4) and this experiment is not intended to make a strong claim on this point. It might

also reflect the amount of information participants were expected to process: The task could likely be made easier by removing the distractor shapes (the squares and triangles) or by presenting children with fewer shapes overall.
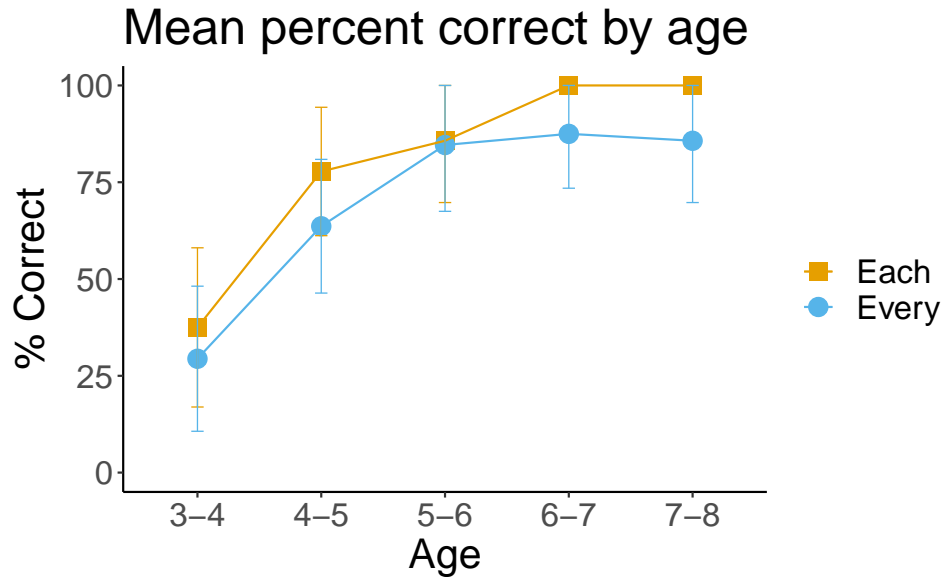
Figure 3.10: Percent correct on the initial "yes"/"no" question (*Is {each/every} circle blue?*), separated by age group. The correct response was always "no".

Only the children who correctly answered the initial question were included in analysis of the center of mass question. For these 109 children, those who first answered the *every*-question recalled the center of mass with greater accuracy than those who first answered the *each*-question ($t = 2.6, p < .05$). Figure 3.11 plots the error distance for each participant (i.e., the distance from their guess to the true center of mass). There was no significant interaction between age and quantifier ($F_{1,105} = .001, p = .971$), suggesting that participants use the predicted verification strategies as early as we can find evidence that they know the meanings of the determiners.

179

Distance from guess to actual center of mass
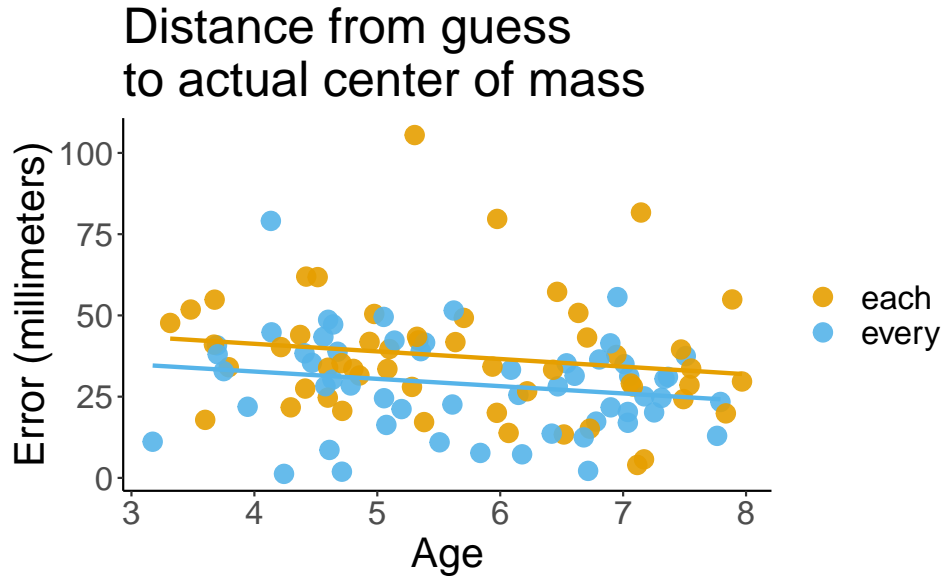
Figure 3.11: Each participant's error distance on the center of mass question (i.e., distance in millimeters from the participant's guess to the true center of mass of the circles). Smaller error distance reflects better memory for the center of mass.

This result is similar to that of Experiment 1 in that *every* leads to superior performance when participants are asked a follow-up question about an ensemble summary statistic. Though the current experiment probes a different summary statistic and uses children instead of adults, it shows a similar advantage for *every* over *each*. One benefit of this task over the other four experiments is that children did not know in advance that they would be asked a follow-up question.

That children show the result as early as they seem to know the meaning of the quantifiers is important. If younger children had correctly responded to the question but not shown the effect, we might be tempted to conclude that the propensity to represent ensembles for *every* and object-files for *each* is a learned strategy or that learners first understand both determiners as universals before understanding exactly which universal concept goes with which pronunciation. The results of Ex-

periment 5 tell against both possibilities. Instead, the lack of interaction between quantifier and age supports the idea that this effect reflects a fundamental detail about these quantifiers' meanings. In particular, if the proposed semantic representations are right, then acquiring the meaning of *each* or *every* is a matter of pairing these representations with the relevant pronunciations.

3.3.4 General discussion

Broadly, the five experiments reported in this section point to the same conclusion. Sentences with *each* bias participants to represent the internal argument of the determiner as a series of independent object-files. As a consequence, they recall individual properties but have degraded knowledge of group properties. Sentences with *every* (and *all*) bias them to abstract away from individuals and represent the internal argument as an ensemble. As a consequence, they recall summary statistics but have degraded knowledge of individual properties.

Importantly, we observe these differences despite the fact that visual displays and truth-conditions were held constant. In the absence of an alternative explanation, details of the semantic representation can be held responsible for the observed variation in verification. In this case, the results are well-explained if *each* is represented in first-order terms, which serve as instructions to create object-files and if *every* is represented in second-order terms, which serve as instructions to create an ensemble.

As always though, there are potential alternative explanations. For example,

what if *each* and *every* (and all quantificational determiners) in fact have second-order representations but *each* moreover has a meaning that entails distributivity? Perhaps this would look something like (143), which is still second-order in quantifying over G, but which also says (in the second conjunct) that each individual must independently satisfy **E**, perhaps owing to the presence of a strong distributivity feature, DIST (Beghelli and Stowell, 1997).

(143) $\iota G(\forall x[Gx \equiv \mathbf{I}x])[\mathbf{E}(G)] \; \& \; \forall y(Gy)[\mathbf{E}y]$ (SO *each* & DIST)

Maybe, given such a representation, participants would forgo forming an ensemble representation of the Gs even though G is quantified into. This is a version of Fodor's challenge from Chapter 1. The idea is that the actual meaning is second-order (or, for Fodor, atomic) and the thing that drives behavior in the task is a separate entailment (or, for Fodor, a strongly associated bit of world knowledge).

But here the use of a distributive predicate in the experiments is crucial. All of the sentences tested used predicates like *be green*. So if the distributivity of *each* were responsible for the results, then the distributivity of the predicates should have led to the same pattern of performance given *every* or *all*.

Put another way, consider the first-order, second-order, and second-order-plus-DIST specifications of *each/every circle is green* in (144).

(144)   a. $\forall x(\mathbf{Circle}(x))[\mathbf{Green}(x)]$ (FO *each*)

           b. $\iota G(\forall x[Gx \equiv \mathbf{Circle}(x)])[\forall y(Gy)[\mathbf{Green}(y)]]$ (SO *every*)

           c. $\iota G(\forall x[Gx \equiv \mathbf{Circle}(x)])[\forall y(Gy)[\mathbf{Green}(y)] \; \&$ (SO *each* &

                $\forall z(Gz)[\mathbf{Green}(z)]$ DIST)

As discussed in section 3.1.2, when a distributive predicate like *be green* takes the place of **E**, it is part of that predicate's meaning that it distributes: $\forall y(Gy)[\textbf{Green}(y)]$. So the distributive conjunct in (144c) is redundant. If its addition causes participants to neglect the instruction to form an ensemble representation of the circles given by "$\iota G(\forall x[Gx \equiv \textbf{Circle}(x)])$", then the same should happen when participants evaluate *every circle is green*, which is represented as in (144b). But this doesn't happen. The distributivity of the predicate does not cause participants to neglect G following *every*. So it seems implausible to think the distributivity of *each* causes them to neglect G. Instead, it seems preferable to say that *each* never implicated G in the first place.

Moreover, in this context it is worth remembering that *every* is often described in the literature as a distributive universal (see section 3.1.2). That is, though *every* is not as strongly distributive as *each*, the two often patten together with respect to distributivity. This may be a pragmatic fact, resulting from the existence of *all*, which takes plural agreement and is preferred for expressing collective thoughts. But given that *each* and *every* often pattern together, if the linking hypothesis were that distributivity triggers object-file representation, it would be strange to then predict that *every* should not trigger object-file representation. That we find *each* and *every* come apart in this respect suggests that the result, at its core, does not reflect distributivity.

Another potential alternative explanation is that these results reflect something about usage, not meaning. There are, of course, plenty of robust usage facts that seem not to matter in experimental contexts. For example, English *most* is

very rarely used in situations where the relevant proportion is close to $\frac{1}{2}$ (Solt, 2016). This usage fact reflects something that speakers know about *most* and it likely influences the felicity of *most*-claims in everyday discourse. But despite this knowledge, participants in experiments like those reported in section 1.3.2 of Chapter 1 readily accept displays in which 55% of the dots are blue as perfectly fine instances of *most of the dots are blue*. So deciding which usage facts should matter is not trivial.

Still, one might wonder if the frequency of *each* and *every* plays a role. The former is likely less frequent in speech (at least, it seems to be slightly less frequent in speech to children; see Chapter 4). If so, processing *each* may, in some sense, take more cognitive effort than processing *every*. That extra effort might have been spent encoding the cardinality or encoding the center of mass, leading to inferior performance following *each*. This might be a concern, if not for the results of Experiments 3 and 4, which showed superior performance following *each*.

Lastly, there is the question of whether existing accounts of *each* and *every* could be said to predict the above results. Tunstall (1998), for example, proposes that *each*, *every*, and *all* differ with respect to the conditions they place on sub-parts of events, which she calls subevents. For instance, if Kermit lifted box$_1$ and put it down, then lifted box$_2$ and box$_3$ simultaneously and put them down, *Kermit lifted the boxes* is said to refer to an event of Kermit lifting the relevant boxes, which is made up of two subevents: the lifting of box$_1$ and the lifting of the other two.

Tunstall proposes that *each* requires each individual in the denotation of its internal argument to be associated with its own subevent, *every* requires only that there be at least two distinct subevents, and *all* has no such requirement on

subevents. This predicts a truth-conditional difference between the sentences in (145). Restricting the domain to the contextually relevant boxes, this proposal predicts that (145a) is TRUE so long as Kermit lifted all of them, (145b) is TRUE just in case Kermit lifted all of them but not all at once, and (145c) is TRUE just in case Kermit lifted all of them but never more than one at a time.

(145)   a. Kermit lifted all the boxes

         b. Kermit lifted every box

         c. Kermit lifted each box

This prediction seems far too strong. But assuming some version of Tunstall's proposal is on the right track, we might wonder if subevent differentiation can account for the above results.

Here, the details of the linking hypothesis are important. The linking hypothesis invoked above – the Interface Transparency Thesis in (135) – maintains that algorithms used during verification will reflect the operations expressed by the representation being evaluated. And one of the main claims here has been that object-files reflect first-order quantification whereas ensembles reflect second-order quantification. By contrast, it is not clear how different amounts of subevent differentiation would be related to either of these two non-linguistic systems. Why would a representation that treats events potentially partially distinct and potentially fully distinct – like Tunstall (1998)'s *every* – trigger an ensemble representation instead of several object-file representations? This is not to say that Tunstall's proposal can't, in principle, account for the above results. However, for it to do so, a new

linking hypothesis would need to be articulated and tested.

Consider one other possible alternative explanation based on an existing account: It is not a particular representational format that gives rise to the present results, but differing presuppositions. In an account aimed at giving a unified treatment of distributivity, aspect, and measurement, Champollion (2017) proposes slightly different presuppositions for different universals. Abstracting away from many details, one one component of the proposal is that when used in sentences like (146a), *each* and *every* come with the presupposition in (146b), whereas *all* comes with the presupposition in (146c).

(146)   a. Each/Every/All the frogs sang

       b. Presupposition of *each/every*: Any singing event e can be divided into one or more singing events whose agents are each atoms

       c. Presupposition of *all*: Any singing e can be divided into one or more singing events whose agents are each small in number compared to the agent of e

Since Champollion does not differentiate *each* and *every* with respect to this presupposition, it cannot, on its own, account for our results.

That said, perhaps we could imagine a slightly different presupposition for *every* that would differentiate it from both *each* and *all*. If *each* is said to presuppose subdivisions into atomic agents, and *all* is said to presuppose subdivisions into small sets of agents, perhaps *every* could be said to presuppose subdivisions into sets of agents each containing a single member. Articulating this post-hoc modification

and spelling out what it would mean for Champollion's proposal and for the linking hypothesis is left for future work.

3.4   Semantic evidence: The genericity asymmetry

In addition to the behavioral evidence, the proposed first-order and second-order representations may also help explain an asymmetry between *each* and the other universals with respect to genericity. Section 3.4.1 describes the contrast and a scope-based proposal for capturing the relevant data. Section 3.4.2 argues for an alternative explanation, which is rooted in the proposed first-order and second-order representations and their associated cognitive systems, ensembles and object-files. The proposed explanation is unorthodox in that it relies on details of the interfacing cognitive systems. If right, the genericity asymmetry with respect to *each* and *every/all* is a distinction that resides in thought, not in grammar.

3.4.1   The data and a scope-based explanation

The relevant facts are that *every* and *all* are compatible with certain generic thoughts whereas *each* seems to resist generic interpretations altogether. As Beghelli and Stowell (1997) put it, a sentence like (147a) can be understood as a claim about dogs in general. The same seems to be true for *all* in (147b).[15] But *each* does not give rise to this sort of interpretation (in this section, I will use "#" to signal the

---

[15] It is worth noting, though, that *all* only has generic import when it takes a bare NP. When it occurs with a definite plural or partitive – e.g., *all (of) the dogs love bacon* – it no longer suggests a generic interpretation. The same is true for *every*: a sentence like *every one of the dogs loves bacon* has no generic interpretation.

lack of any generic interpretation). Instead, the *each*-version in (147c) seems to be a claim about contextually salient dogs.[16]

(147)   a. Every dog loves bacon

        b. All dogs love bacon

        c. #Each dog loves bacon

Beghelli and Stowell (1997) also discuss examples like those in (148) and (149), which they attribute to Gil (1992).

(148) After a lifetime of investigation, Suzie came to a striking discovery:

        a. Every language has over twenty color words

        b. All languages have over twenty color words

        c. #Each language has over twenty color words

(149) Suzie just discovered four new languages and interestingly,

        a. #Every language has over twenty color words

        b. #All languages have over twenty color words

        c. Each language has over twenty color words

The observation is that *every* and *all* in some sense allow for a larger domain of quantification. In (148), they both are consistent with the intended interpretation

---

[16] Italian, which also has three universal quantifiers, gives rise to similar judgments: *ogni* and *tutti* allow for this sort of generic interpretation, whereas *ciascun* does not (Nicolò Cesena-Arlotti, pc and Martina Abbondanza, pc). A series of studies currently underway is exploring the possibility that *ogni*, *tutti*, and *ciascun* also pattern together with respect to English *every*, *all*, and *each* in the sorts of experiments presented in section 3.3 (in collaboration with Martina Abbondanza, Florian Schwarz, Francesca Foppolo, Marco Marelli, and Jeremy Zehr).

that Suzie only studied a handful of languages but feels confident enough to generalize beyond them. What makes the *each* version anomalous is that her striking discovery generalizes no further than the particular languages she studied (making it not so striking after all). Replacing *striking discovery* with *universal generalization* may help to bring out the same intuition. A universal generalization can be made over *every language* or *all languages*, but not over *each language*.

On the other hand, *every* and *all* can sometimes lead to anomaly when the claim is meant to be restricted to a very local and contextually-determined domain, like the four new languages in (149). Using *every* or *all* gives the sense that the speaker is generalizing far too broadly given Suzie's discovery. In this case though, *each* is perfectly at home.

To capture this distinction, Beghelli and Stowell (1997) appeal to a proposed difference in the scope positions of *each* and *every*, driven by feature-checking movement. The former is said to have a strong distributive feature, DIST, causing it to move to the specifier of a distributive projection, which is situated relatively high in the tree. The latter is said to only have a weak distributive feature, allowing it to optionally remain lower in the tree.

Then, taking inspiration from Szabolcsi (1997), they assume that DPs headed by *each* or *every* introduce discourse referents in the form of variables that range over sets. These "set variables" can be bound by other operators (though not by the distributive operator, as they note: "this operator applies at a different level, that of the elements of the set" (p.17)). Generic contexts like (148) are argued to result from the presence of a covert generic operator, GEN. Putting all the elements

together: A DP like *every language* introduces a set variable that can optionally get bound by GEN, but a DP like *each language* must move above GEN to check its strong DIST feature, and its set variable thus cannot be bound by GEN.

Some of the theoretical machinery invoked in the above proposal is independently supported. The notion of a set variable, for instance, is borrowed from Discourse Representation Theory (though it should be noted that the experimental results in section 3.3 potentially tell against the idea that *each* introduces a set variable). And the idea of features triggering movement is not uncommon. But the particular features proposed (e.g., strong and weak DIST) and the functional projections targeted for movement (e.g., DistP) are largely sui generis (see Larson (2021) on the explanatory adequacy of cartographic approaches). The same can be said about the generic operator GEN, whose semantics has been notoriously difficult to cash out in a way that will capture the wide range of data that falls under the header of "genericity" (see Leslie and Lerner (2016) for a helpful review).

To be sure, the genericity asymmetry represents only a fragment of the data that Beghelli and Stowell's scope-based approach aims to capture. For example, movement to the specifier of a distributive projection is also proposed to account for the strong distributivity of *each*, discussed in Chapter 1. That is, DistP is said to house the distributive operator with which *each* always associates (and with which *every* only optionally associates).

Surányi (2003) brings up a number of problem with this analysis, including the fact that *each* gives rise to extra-clausal distributive readings, whereas *every* appears to be clause-bounded. For example, in cases like (150) *every* can give rise

only to a single yes/no answer whereas *each* can give rise to a pair-list response, seemingly scoping outside of a whether island. And in (151), an example from Szabolcsi (2010), *every* cannot occur but *each* seemingly scopes outside of a relative clause island.

(150) Determine whether {each/every} number in this list is odd

(151) A timeline poster should list the different ages/periods (Triassic, Jurassic, etc.) and some of the dinosaurs or other animals/bacteria that lived in {each/ *every one}

This distinction is not predicted given Beghelli and Stowell's taxonomy and it is not immediately clear how it should be accommodated. As a solution, Szabolcsi floats the following idea: "Perhaps the distributivity operator is part of the lexical semantics of *each*, in contrast to *every*" (p.127). This is in line with the current proposal that *each* has a first-order representation.

Beghelli and Stowell (1997) also aim to explain other scope preferences. For example, the relationship between *each/every* and negation is discussed and contrasts like (152) are captured.

(152) Kermit didn't eat {??each/every} apple   (with non-focused intonation)

It is worth noting, though, that this judgement is not universally agreed upon, and Beghelli and Stowell acknowledge that it "depart[s] from what is generally assumed about such data" (p.27).

Gagnon and Wellwood (2011) extend their proposal to cover an interesting asymmetry with respect to epistemic containment. Namely, *every* seems unable take scope over epistemic modals like *might* in (153a), but *each* can (both can take scope over root modals like *could*, as in (153b)).

(153) a. {Each/??Every} student in this room might$_{epistemic}$ be the smartest (as far as I know)

b. {Each/Every} student in this room could$_{ability}$ be the smartest (if they studied)

Given independent evidence that epistemic modals scope relatively high compared to root modals, and independent evidence that modals can bind free variables, this data seems to fit nicely with a taxonomic view.

Of course, as Gagnon and Wellwood note, the strict scope taxonomy proposed in Beghelli and Stowell (1997) predicts a number of scope orderings to be impossible, and these predictions may not be borne out. For example, Beghelli and Stowell predict that "A CQP [e.g., *fewer than four frogs*] in object position should never be able to take inverse scope over a GQP [e.g., *a frog*] or DQP [e.g., *each frog*] occurring in subject position" (p.11). Brendel (2019) reports experimental evidence against this strong prediction. And considering the subtlety of other judgments used to motivate the view – like (152), for instance – additional controlled judgement studies along these lines are necessary.

In any case, the alternative account of the genericity asymmetry proposed below is not intended to explain all of the data that has been captured on the

taxonomic view or argue against a hierarchy of fixed scope positions for different determiners. The aim is more modest: to suggest that the genericity asymmetry between *each* and *every* does not arise because of a scope difference with respect to GEN. It is left for future work to determine whether the first-order/second-order distinction plays a role in explaining other apparent facts captured by appealing to different scope positions. But it is worth noting that the claim that different determiners have different scope positions and the claim that different determiners have different representational formats are not mutually exclusive. Following Szabolcsi (2010)'s suggestion, it might ultimately be best to combine the two views.

3.4.2   An extralinguistic alternative

The crux of the idea is this: The second-order meaning of *every* (and *all*) is an instruction to form an ensemble representation, and ensembles have properties that make them compatible with the sort of generic thoughts at issue here. On the other hand, *each*'s first-order meaning is an instruction to form object-file representations, and object-files do not support generic thoughts.

To be clear, the claim is not that *every* in any sense has a "generic meaning." Both *each* and *every* express universal generalizations. The difference is in how they express that generalization (in first-order or second-order terms) and, crucially, in what cognitive systems are implicated as a consequence. The first-order *each* represents the universal generalization in a way that promotes object-file thoughts, which do not support projecting beyond the local domain. The second-order *every*

represents the same generalization in a way that promotes ensemble thoughts, which do support projecting beyond it.

As observed above, the sorts of generics at issue here are ones that require projecting beyond the local domain. What goes wrong with *each* in (148) is that it seems to be about the local domain (i.e., the languages that Suzie studied) and doesn't allow for predictions about new instances. The versions with *every* or *all*, in contrast, support predictions about new languages: They too will have over twenty color words. This same distinction arises with examples like (154), where *each* leads to a far weaker claim about gravity acting on some particular objects.

(154)   a. Gravity acts on every object

        b. Gravity acts on all objects

        c. #Gravity acts on each object

But there are other cases that fall under the header of "genericity" for which the *each*/*every* asymmetry does not arise. The *each*-, *every*-, and *all*-versions of (155), for example, all seem equally FALSE, whereas a bare plural NP can be understood generically given the same predicate, as in (155a).

(155)   a. Ticks carry Lyme disease

        b. #Each tick carries Lyme disease

        c. #Every tick carries Lyme disease

        d. #All ticks carry Lyme disease

This sort of example presents an issue for Beghelli and Stowell (1997)'s account. If GEN is responsible for the generic interpretation of (155a), then what precludes it from binding *every*'s set variable in (155c) and likewise giving rise to a generic interpretation?

The present account is that the genericity asymmetry between *every/all* and *each* arises because ensembles, but not object-files, support projecting beyond the local domain. Projection is not at issue in (155), attributing a property to a kind is. However, if the example is rephrased, the asymmetry seems to return:

(156)   a. #Each tick can carry Lyme disease

        b. Every tick can carry Lyme disease

        c. All ticks can carry Lyme disease

In (156), the property of being able to carry Lyme disease is attributed to ticks as opposed to the natural kind TICK. So the relevant fact to explain seems to be that *every* supports projecting beyond the local domain whereas *each* does not. The proposal is that this difference results from details of the cognitive systems related to their first-order or second-order representations.

Why do ensembles, but not object-files, support projecting beyond the local domain? As discussed in section 3.2.2, ensembles represent a collection of individuals by abstracting away from the individuals themselves and encoding summary statistics. In other words, ensemble representations support generalization by abstracting away from individual instances. On the other hand, as discussed in section 3.2.1, object-files are first and foremost indices. The properties bound to object-files

are secondary. What makes an object-file an object-file is not its color or shape or size. Representations created by this system do not, on their own, support generalization. That two object-files share some properties is an accident as far as the object-file system is concerned.

These differences suggest a difference in these two systems' abilities to license predictions about new items (i.e., to project). Imagine being presented with four green circles. If you represent these circles as an ensemble, you will have extracted their average hue and the range of hues. This naturally licenses a prediction about the next circle you'll encounter (assuming it will also be a member of the current ensemble). On the other hand, if you formed four object-file representations, you will have encoded the individual circles' spatial positions and bound each individual hue to the relevant index. You might be able to predict that the next circle would also be green. But to do so, you would have to rely on separate systems (e.g., pattern recognition or a process of induction). This prediction would not come for free by virtue of representing independent object-files.

To be sure, this difference with respect to projecting beyond the local domain awaits empirical support. It does make a number of testable predictions. For one thing, if participants are encouraged to represent some objects as an ensemble, they should be more willing to make a prediction about a new object than if they are encouraged to represent the very same objects as independent object-files. This difference should, by hypothesis, be present whether participants are encouraged to represent the objects in a particular way for linguistic reasons (e.g., because *each* or *every* was used) or non-linguistic reasons (e.g., because the level of homogeneity

in the visual display was small or large). Moreover, when participants make a prediction in the object-file case, they should do so not on the basis of summary statistics but on the basis of pattern recognition. For example, if three circles – one big, one medium, and one small – are represented as a series of object-files, participants might expect the next circle to be tiny, in keeping with the pattern. But representing them as an ensemble instead suggests using the average size for setting future expectations. These empirical predictions will be tested in future work.

For the time being though, assume that ensembles, but not object-files, support projection. Connecting back to the genericity asymmetry, the idea is that the *each* and *every* sentences in (148) have the truth-conditionally equivalent semantic representations in (157).

(157)   a. Each language has over twenty color words

   $$\forall x(\textbf{language}(x))[\textbf{has-over-20-color-words}(x)]$$

   b. Every language has over twenty color words

   $$\iota G(\forall x[Gx \equiv \textbf{language}(x)])[\textbf{has-over-20-color-words}(G)]$$

The reason that (157b) is compatible with a generic interpretation is that its semantic representation serves as an instruction to create an ensemble representation of the languages. This representation, call it G, supports projecting beyond the local domain in the way discussed above. In particular, if the predicate **has-over-20-color-words** applies to G, a prediction about new members of G is licensed for free because G is an ensemble representation. If a new language is one of the Gs, it will

197

likely share properties with the other Gs.

On the other hand, the semantic representation in (157a) provides an instruction to consider the languages as independent object-files. These representations do not support predictions about new languages, at least not for free. A prediction that the next language should be similar would have to be the result of induction (e.g., the predicate held of language 1, language 2, and language 3, ... maybe it holds of every language).

In sum, the asymmetry between *each* and *every/all* in examples like (148) and (149) may owe its existence more to the details of non-linguistic cognition than to the details of grammar. That is, while details about their meanings explain the link between the quantifiers and ensembles or object-files, it is the details of those non-linguistic representations that explain why only *every* and *all* are compatible with certain generic interpretations. The cognitive system called upon by *every* and *all* supports the relevant type of generic thought, but the cognitive system called upon by *each* does not.

This correctly predicts that the asymmetry between *each* and *every/all* arises only when projecting beyond the local domain is at issue. It does not arise when, for example, properties are attributed to kinds (as natural kind concepts are not ensemble representations). So this account does not suggest that all cases of genericity will be explained by appeal to ensembles and object-files. But it does reduce the explanatory task of other theories of genericity slightly: They need not explain why certain quantificational determiners are compatible with certain generic interpretations. Given that overtly quantified sentences are rarely the main target of inquiry

of work on genericity, this may come as a welcome conclusion.

3.5   Chapter summary

This chapter provided some reasons for thinking that *each* has a first-order semantic representation along the lines of (158) whereas *every* has a second-order representation along the lines of (159).

(158) $\lambda\mathbf{I}.\lambda\mathbf{E}.\forall x(\mathbf{I}x)[\mathbf{E}x]$ (FO restricted, 0G)

(159) $\lambda\mathbf{I}.\lambda\mathbf{E}.\iota G(\forall x[Gx \equiv \mathbf{I}x])[\mathbf{E}(G)]$ (SO restricted, 1G)

As discussed in section 3.1, the difference between these specifications is, in the first place, a matter of logical syntax: First-order representations allow quantification into lowercase variable positions (e.g., "$\forall x$") whereas second-order representations also allow quantification into uppercase variable positions (e.g., "ιG"). So while (158) implicates only individual things – each x – that meet the conditions supplied by the internal and external arguments, (159) also implicates a group – the Gs – made up of all the things that meet the internal condition.

Section 3.1.2 showed that, at least for the basic cases, these representations are compatible with the standard approach to distributivity. The proposed first-order representation only allows distributive interpretations, whereas the proposed second-order representation makes available both distributive and collective interpretations. Distributivity in the second-order case is achieved either by a covert distributivity operator (e.g., for ambiguous predicates like *sing happy birthday*) or by a distributive meaning postulate (e.g., for predicates like *smile*).

This logical distinction between first-order and second-order representations neatly corresponds to the well-established psychological distinction between object-files and ensembles, discussed in section 3.2. Like first-order variables, object-files are pointers to individuals (to which properties like color or size may be bound). Like second-order variables, ensembles can be thought of as pointers to many individuals at once. Given working memory limitations, this is cognitively achieved by abstracting away from the individuals and encoding the group in terms of summary statistics like cardinality, center of mass, and average size.

Leveraging this link, the experiments in section 3.3 find that participants have better memory for ensemble summary statistics (cardinality, center of mass) after evaluating *every*-sentences than after evaluating truth-conditionally equivalent *each*-sentences. On the other hand, they have better memory for individual properties (hue) after evaluating *each*-sentences. These results suggest that, from a young age, *every* encourages representation of ensembles whereas *each* encourages representation of object-files. Given a link between the psychological and the logical representations, these results are well-explained by the proposed first-order and second-order representations.

Finally, section 3.4 proposed a novel account of judgments pertaining to genericity. In particular, while *every* can be used to project beyond the local domain in sentences like *every dog loves bacon*, *each* does not as naturally give rise to the same generic thought. As opposed to situating this contrast in the grammar (e.g., scope positions relative to a covert generic operator), the difference was proposed to reside in the related cognitive systems. In particular, ensemble representations

support projecting beyond the local domain, whereas object-files do not. Though this analysis requires further empirical support, it could potentially explain the genericity asymmetry as a consequence of non-linguistic cognition.

Chapter 4: Acquiring *each* and *every*

If the preceding chapters are on the right track, they suggest that knowing the meaning of a quantificational determiner consists in representing its (truth-conditional) content in a particular format. For *each*, this means representing a restricted first-order universal concept, like (160a), and for *every*, it means representing a restricted second-order universal concept, like (160b).

(160)   a. $\forall x (\mathbf{I}x)[\mathbf{E}x]$ *each*

   b. $\iota G(\forall x[Gx \equiv \mathbf{I}x])[\mathbf{E}(G)]$ *every*

This naturally raises an acquisition question: What leads learners to pair one pronunciation ("each") with one universal concept and another pronunciation ("every") with logically-equivalent but subtly different universal concept?

This acquisition question only gets more complicated when *all* is added to the mix. For simplicity though, this chapter will focus on *each* and *every* and leave *all* for future work. One motivation for doing so is their grammatical similarity (e.g., they both require singular agreement). Another reason for focusing first on *each* and *every* is that they occur in child-ambient speech at relatively the same rate, which is an order of magnitude less often than *all*. For example, in the corpus analysis reported in section 4.1, *each* occurred 542 times in a sample of over 1.7

million child-ambient utterances, *every* occurred 768 times, but *all* occurred 20,558 times.

Assuming that children hear between 900,000 and 2.5 million utterances per year (Hart and Risley, 1995, 2003), they likely hear *each* somewhere between 286 and 794 times per year, *every* somewhere between 405 and 1,125 times per year, and *all* somewhere between 10,843 and 30,119 times per year. For comparison, consider how often children hear *mommy*, which 98% of twenty-four-month-olds produce according to Wordbank (Frank et al., 2017). Based on an estimate from the same corpora used here, children hear *mommy* between 5,942 and 16,505 times per year (i.e., less frequently than *all*). On the other hand, *each* and *every* are on a par with the verb *pour*, which children likely hear between 407 and 1,130 times per year and which only 25% of twenty-four-month-olds produce.

The high frequency with which learners encounter *all* does not answer the question of how they infer its meaning. But it does mean they have more opportunities to do so. On the other hand, given their relative infrequency, it is no surprise that learners are often said to achieve an adult-like understanding of *each* and *every* relatively late in development. For one thing, children generally start to produce *each* and *every* later than they start to produce *all*, as seen in Figure 4.1.

To be sure, production generally follows comprehension, making it only a rough tool for pinpointing age of acquisition. But in terms of comprehension, many studies that present four- to eight-year-old children with universally quantified sentences report failures of various sorts (e.g., Donaldson and Lloyd 1974; Philip 1994; Brooks and Braine 1996; Musolino et al. 2000; Drozd 2001; Syrett and Musolino 2013;
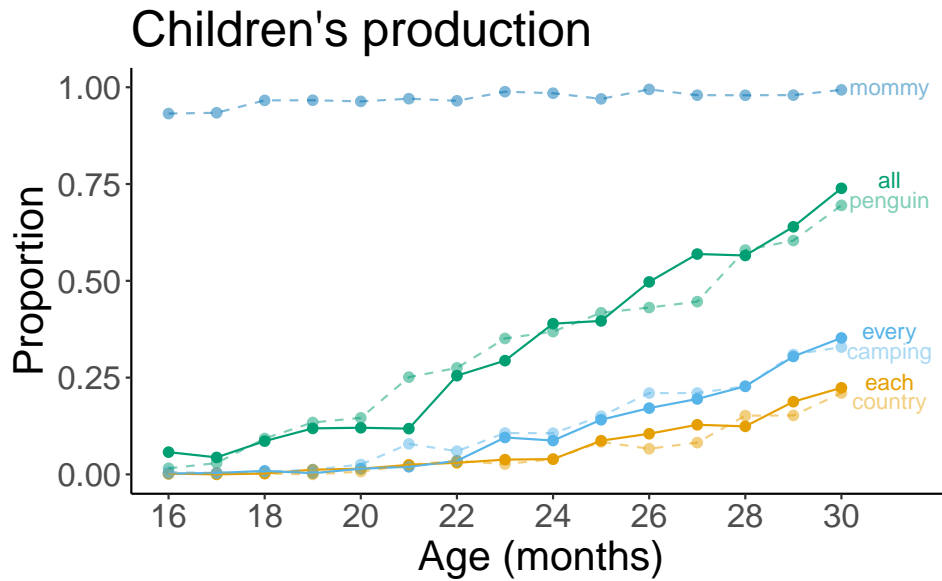
Figure 4.1: Proportion of 16- to 30-month-olds producing *each*, *every*, and *all*. Words with similar production trajectories are included for comparison. Data from Wordbank (Frank et al., 2017).

Achimova et al. 2017).[1] Of course, some accounts argue that these failures reflect performance limitations and task effects (Crain et al. 1996; Musolino and Lidz 2006; Minai et al. 2012). But even in the relatively simple (albeit one-trial) task reported in section 3.3.3 of the previous chapter, children below age five were below 75% correct.

This chapter will not contribute to the debate over what non-adult-like performance in certain experimental situations reflects. As a result, it will not make

---

[1] "Quantifier spreading" errors have been of particular interest, going back to Inhelder and Piaget (1958). In a typical quantifier spreading task, a child might be shown four umbrellas, three of which are held by turtles, and asked whether they think *every turtle has an umbrella*. They often incorrectly answer "no," as if they understand the sentence to mean *every umbrella is being held by a turtle.* These results are robust and have been well-replicated. Some accounts attribute these errors to non-adult-like understandings of e.g., *every* (Philip 1994; Drozd 2001; Roeper et al. 2011). But there are also reasons for thinking that these results reflect pragmatic details of the task and not an underlying difficulty understanding (Crain et al., 1996). For example, if there are three un-held umbrellas instead of just one in the above case, children are more likely to correctly respond "yes" (Sugisaki and Isobe 2001; Minai et al. 2012).

any claims about the age of acquisition of the universal quantifiers. Instead, its aim will be to get the the "how" question on the table – what information do learners use to infer the correct meanings for *each* and *every*? – and to offer some initial evaluations of possible directions.

In a sense, identifying the target of learning with a particular representational format makes this question more difficult than it would be if the target of learning were particular truth-conditions, represented in any format. If the target of learning were universal truth-conditions, independent of format, then learners might plausibly notice that whenever parents use *each* or *every*, they exhaustively predicate something of some domain. This is not to say acquiring the right truth-conditions would be easy. Any time it's TRUE that *every **I** is **E***, it's also TRUE that, for example, *some **I** is **E***. So more needs to be said about why learners don't settle on the hypothesis that *every* means *some* (see Piantadosi et al. (2008) and Rasin and Aravind (2021) for potential solutions to this "subset problem").

But even assuming a solution to the challenge of how to acquire universal truth-conditions, the further question of how to acquire the right representational format remains. Since (160a) and (160b) are truth-conditionally equivalent to one another and to many other ways of specifying universal content, no amount of TRUE *each*- or *every*-claims will allow the learner to decide between the possibilities. Something more subtle in their input must signal the difference between *each* and *every*.

Section 4.1 considers three possibilities, based on three linguistic phenomena: (i) that *every* but not *each* can occur with certain collective predicates, (ii) that *each* but not *every* gives rise to pair-list readings, and (iii) that *every* but not *each*

encourages projecting beyond the local domain. While information about (i) and (ii) are not robustly available in child-ambient speech, information about (iii) seems to be. That is, parents use *every*, but not *each*, to make broad generalizations that project beyond the local domain. This difference in usage leads to lower-level concomitants, including a large difference in what routinely gets quantified over. Section 4.2 spells out one proposal for how learners might use these low-level differences to infer the correct meanings. And section 4.3 discusses future work that will test each element of this proposal.

4.1 How parents use universal quantifiers: Corpus findings

Data from this corpus investigation – first reported in Knowlton and Lidz (2021) – consists of utterances containing *each* and *every* from a segment of the North American English portion of CHILDES. This sample includes both child-directed and overheard speech, so the broader term "child-ambient" is used, where appropriate. Only corpora that contained speech to typically-developing children under eight were included (see Appendix A for the list of corpora used and associated citations).

Setting the cutoff at eight-years-old assumes that the resolution to the age of acquisition question raised above will not be that children acquire adult-like competence with *each* and *every* at, say, age nine. But it may turn out that children know the meanings of these determiners far earlier than eight-years-old, in which case our age range extends too high. Just as importantly, our age range likely also

extends too low. The youngest child in this sample is 7-months-old, and there is no guarantee a that 7-month-old is developmentally equipped to begin acquiring *each* and *every*. This initial corpus investigation consequently idealizes away from the role of development in language acquisition (Perkins, 2019). But given how infrequently *each* and *every* appear in child-ambient speech, age constraints were kept loose to include as many utterances as possible.

For similar reasons, cases of *each* in non-canonical determiner positions were not excluded. These include examples like (161a) from the Soderstrom corpus (Soderstrom et al., 2008) and (161b) from the Warren corpus (Warren-Leubecker and Bohannon, 1984) (the age in parentheses is the age of the child at the time of their parent's utterance).

(161)   a. No, the pillows each stay where they belong, please   (0;08)

          b. We got thirty [points] each!   (6;02)

There is debate over whether "floating *each*" examples like (161a) are cases of a determiner being stranded after movement (Sportiche 1988; Benmamoun 1999) or an adverbial modifier that never forms a constituent with a noun phrase (Dowty and Brodie 1984; LaTerza 2014). And "binomial *each*" examples like (161b) are more clearly cases of adverbial quantification (Safir and Stowell, 1987). So including these instances of *each* – which together make up approximately 13% of all *each* utterances, as seen in Table 4.1 – may not be an innocent choice. But assuming learners treat determiner and adverbial *each* as univocal (as opposed to treating them as homophonous or polysemous), it's reasonable to assume that children will

207

use these non-canonical instances as evidence about *each*'s meaning.

	Total	Determiner	Floating	Binomial	Unknown
each	542	465	61	12	4

Table 4.1: Occurrences of determiner and adverbial *each* as standalone lexical items in child-ambient speech.

In fact, given that *every* does not float or get used binomially, the fact that *each* does might provide some evidence to learners as to its meaning. Maybe learners reason that floating quantifiers are likely to signal distributivity, perhaps by noticing that the parent uttering (161a) put the pillows back in their respective positions one at a time. Consequently, maybe learners can then infer that *each* must have a first-order meaning (since a first-order representation would enforce distributivity, as discussed in Chapter 3).

On the other hand, *all* can be used in non-canonical positions (e.g., *the pillows all stay where they belong, please*). And by hypothesis, *all* and *every* are alike in having restricted second-order representations (though likely not the same representations). So if floating and binomial *each* are crucial pieces of evidence that push learners toward a first-order representation, more would have to be said about why they don't likewise push learners in that direction for *all*. For now, the role of *each* in non-canonical positions will be set aside.

There were some constraints on the data considered. Children's production was not included in this sample; Only parents' utterances were examined. Moreover, only cases where *each* and *every* were transcribed as standalone items were considered (e.g., lexicalized quantifiers like *everybody* and *everyone* were not included).

As noted above, this yielded 542 instances of *each* and 768 instances of *every*.

We are now in a position to ask: Are the differences in the idealized distribution identified by linguists also reflected in speech to children? In particular, the sections below consider the *each/every* asymmetries with respect to distributivity and genericity, both of which were discussed in Chapter 3. If these differences are present in child-ambient speech, then they could potentially be used to drive learning. But if they aren't, they won't play a role in acquisition.

4.1.1 Collective predicates and pair-list readings

As discussed in section 3.1.2, *each* is strongly distributive. One consequence of its distributivity is that a quantificational DP headed by *each* is unable to combine with certain collective predicates – like *lifted the piano together* or *gathered in the hall* – that are able to combine with *every*, as seen in (162).

(162)   a. {Every/#Each} student lifted the piano together

        b. {Every/#Each} student gathered in the hall

This suggests a natural usage difference between *each* and *every* that, if present in a sufficient quantity, could be informative for learners.

But in the 1,310 *each*- and *every*-utterances considered, there was only a single unambiguously collective predicate: (163) from the HSLLD corpus (Dickinson and Tabors, 2001).

(163) Don't we watch it like every week together?   (4;11)

And in this case, *every* is not the head of a subject DP that combines with the collective predicate *watch it together.* Instead, *every* is part of a temporal adjunct that would have allowed *each* as well (e.g., *Didn't we watch that show together each night of vacation?*). This makes it implausible that ability to appear with collective predicates explains how children acquire *every.*[2]

Still, *each*'s strong distributivity surfaces in other ways. One such way is that it supports pair-list responses to questions even in cases where *every* does not, as in (164) (Beghelli, 1997).

(164) a. Which book did you loan to each student?

Aspects to Annie, *Syntactic Structures* to Shirley, and *SPE* to Pierce

b. Which book did you loan to every student?

#*Aspects* to Annie, *Syntactic Structures* to Shirley, and *SPE* to Pierce

Of course, for facts like this to be useful, learners would need to be exposed not only to the question, but to the corresponding answer (e.g., perhaps in overhead conversation between two adults).

This possibility is unlikely. In our sample, there were only 11 instances of *each* co-occuring with a WH-question and only 19 instances of *every* co-occurring with a WH-question. Of these, only two utterances were questions that could have plausibly received a pair-list response and both were directed toward a child. The relevant examples are (165a), from the Snow corpus (MacWhinney and Snow, 1990), and (165b), from the HSLLD corpus (Dickinson and Tabors, 2001).

_____

[2] It may be more useful when it comes to *all.* Preliminary coding of *all*-utterances reveals a few cases of collective predicates – e.g., *let's all sing a song together* – though these are still rare (out of 1,000 randomly sampled utterances, there were 13 collective predicates).

(165)   a. Father: What do you think each animal is about to do?

Child (3;04): Clean up that mess

b. Mother: What did you play every day while you were there?

Child (4;11): ...the water game

In response to (165a), a pair-list response could have been given (e.g., "the tiger is about to hunt, the zebra is about to run, and the frog is about to laugh"). The same could be said of (165b), if not for the use of *every* instead of *each*. In neither case did the child offer a pair-list response (and in the former case, the father did not make them aware of this possibility after the fact). Therefore, pair-list answers to *each*-questions cannot be the critical information that children use to learn its meaning (at least insofar as the corpus is representative in this respect).

These are not the only imaginable correlates of the distributivity asymmetry. Scope may also play a role, given that adults generally prefer *each* to take wide scope more often than other quantifiers in scopally-ambiguous sentences (Ioup 1975; Kurtzman and MacDonald 1993; Feiman and Snedeker 2016). On the other hand, children often opt for surface-scope interpretations of such sentences, and seem to have difficulty revising this initial commitment (Lidz and Musolino 2002; Musolino and Lidz 2003). So useful learning instances will likely consist in cases where *each* takes wide scope and appears before other operators on the surface. Determining the prevalence of these cases is left for future work. In the meantime, the present results suggest that *each*'s strong distributivity may not be what differentiates it from *every* in child-ambient speech.

4.1.2   Encouraging projecting beyond the local domain

Another possibility is that the genericity asymmetry discussed in section 3.4 of the previous chapter drives differences in how parents use *each* and *every*. In particular, they may use *every* to convey broad generalizations that project beyond the local domain and use *each* to generalize only over the local domain. Whether an utterance was intended to convey a generic thought of this sort is an abstract notion that can be difficult to judge. Moreover, learners do not have direct access to their parents' intention to produce a generic thought. For both of these reasons, features that might reflect detectable footprints of genericity were coded instead of genericity itself. The particular lower-level differences considered are in (166).

(166)   a. Quantifying over times vs. individuals

b. Explicitly vs. contextually restricting the domain

c. Occurring in a clause in present tense vs. other tenses

d. Appearing in an adjunct vs. in a verb's argument

Subsequent sections will discuss why these lower-level differences plausibly reflect the higher-level difference between one quantificational determiner (*every*) being used to convey a thought that projects beyond the local domain and another (*each*) being used to generalize over a local domain.

But first, it may be useful to consider some examples of parents using *every* to get across thoughts that are meant to project beyond the local domain. To take one example, consider (167), which is found in the Peters/Wilson corpus (Wilson

and Peters, 1988).

(167) Father: Do you want a cookie?

Seth (1;05): No

Father: Every time I give you one, you throw it

Presumably, Seth's father was not intending to highlight each past event of cookie-giving but to make a larger claim about what happens, in general, when he gives his son a cookie. This generalization projects beyond the local domain in the sense that what it quantifies over – events of cookie-givings – are not present at the time of utterance. And moreover, it licenses a prediction about the future: Seth will throw the next cookie he receives. This is exactly the sort of generic generalization that *every* naturally supports, but *each* resists.

Similar examples are common; consider those in (168) (the first two are also from the Peters/Wilson corpus and the last two are from the Weist corpus (Weist and Zevenbergen, 2008)).

(168) a. You turn into a wild man every time we get out (1;10)

b. I'm just tired of fighting with you about every damn thing (2;00)

c. Every time I ask a question, you say you don't know (2;10)

d. Don't you draw like every day pretty much? (4;09)

These utterances seem to project beyond the local domain in the same way as (167). They quantify over things that are certainly not present (events of going

213

out, possible things to fight about, events of asking questions, days) and they license predictions (the child will misbehave, the next thing will be fought about, the child will say they don't know the answer to the next question, the child will draw tomorrow). Compare these to some example *each*-utterances drawn from the same two corpora:

(169)   a. How are you gonna walk when you got a Teddy in each hand?   (1;06)

      b. In fact you have five fingers and five spiders; you could put one spider on each finger   (2;04)

      c. Is it one dollar for all nineteen lizards or one dollar for each lizard? (4;05)

      d. Well tell me about each one, so I can decide which one   (4;07)

The utterances in (169) appear to express generalizations about very local domains (the child's hands, the child's fingers, toy lizards in a shopping game, toy dinosaurs in a shopping game). And they don't license predictions about anything beyond what's immediately being quantified over.

The next four sections report on the potential distributional footprints of genericity in (166). Having found these features in child-ambient speech, sections 4.2 and 4.3 discuss how and whether children could use them in acquisition.

4.1.3   Quantifying over individuals vs. times

Utterances were coded for the types of things that were quantified over. Examples for each category are given in (170).

(170)   a. **Individuals**: *each toy*; *every painting*

   b. **Times**: *every year*; *every time*

   c. **Locations**: *every place we went*; *each corner of the room*

   d. **Degrees**: *all clean*; *every little bit*

   e. **Events**: *every move you make*; *every step you take*

As seen in Table 4.2, parents most often use *each* to quantify over individuals (e.g., *you need to eat each piece of broccoli if you want dessert!*), whereas they most often use *every* to quantify over times (e.g., *every time we have broccoli, you leave leftovers!*). Statistically, *every* is more likely to be used to quantify over times than *each* ($\chi^2 = 264.6, p < .001$) and *each* is more likely to be used to quantify over individuals than *every* ($\chi^2 = 150.7, p < .001$).[3]

Det	Total	Individuals	Times	Locations	Degrees	Events	N/A
each	542	**406**	32	102	0	0	2
every	768	212	**486**	28	10	20	12

Table 4.2: What gets quantified over in child-ambient speech. Particularly relevant values are in bold.

Why should this robust difference in what gets quantified over reflect the genericity asymmetry discussed above? If parents do use *each* to generalize only over local domains, it should come as no surprise that they mostly quantify over individuals. Toys, drawings, fingers, and the like are more likely to be in the local

---

[3] Expected values were calculated using the rate of quantifying over individuals and times in parents' utterances with either *each* or *every*. Future work will use rate of quantifying over a particular type in quantified statements more generally.

domain – and, moreover, to be physically present at the time of utterance – than, for example, events of dinner-eating.

Of course, individuals quantified over need not be in the local domain. Parents can and do use *every* to make projectable generalizations about individuals (e.g., *every painting you do is that color*). But they more often use *every* to generalize over times. Perhaps this is because situations are more useful things to generalize over in a way that projects beyond the local domain. This would not reflect a fact about language, but about what sorts of things parents want to communicate to their children. A phrase like *every raven is black* is perhaps the canonical example of a quantified expression with a generic interpretation (Hempel, 1937, 1945). But it doesn't seem to be the sort of thing that parents are likely to use *every* to express. And given that parents' preference leads to a large low-level difference in what gets quantified over linguistically, perhaps it is pedagogically useful in an unintended way (i.e., for signaling the difference between *each* and *every*).

4.1.4   Explicit restriction with a relative clause

Since parents often quantify over times, and since *every time* without further qualification picks out a large swath of situations, we might expect them to explicitly restrict the domain (e.g., *you turn into a wild man every time we get out*). To this end, utterances were coded for whether the NP being quantified over was modified with a relative clause (e.g., *every time (that) we get out* as opposed to just *every time*). As seen in Table 4.3, occurrences of *Q NP* by and large are not modified by

a relative clause. However, parents are more likely to use a relative clause to modify NPs quantified over by *every* than by *each* ($\chi^2 = 80.6, p < .001$).

Det	Total	No	Yes	N/A
each	542	**530**	11	1
every	768	604	**152**	12

Table 4.3: Number of times a relative clause is used to modify the quantifier phrase, explicitly restricting the domain of quantification. Particularly relevant values are in bold.

Perhaps surprisingly, this result is not solely driven by the fact that parents most often quantify over times with *every*. The result holds even when considering just cases in which individuals are quantified over: Parents are more likely to modify *every* $NP_{individual}$ with a relative clause than *each* $NP_{individual}$ ($\chi^2 = 43.8, p < .001$; Table 4.4).

QP	Total	No	Yes	N/A
each $NP_{individual}$	406	**399**	6	1
every $NP_{individual}$	212	177	**33**	2

Table 4.4: Restricted to cases of quantifying over individuals, number of times a relative clause is used to modify the quantifier phrase, explicitly restricting the domain of quantification. Particularly relevant values are in bold.

One possible explanation for this difference is that when parents use *each*, any intended domain restriction is already contextually salient. This would make explicit restriction with a relative clause unnecessary. For example, in (171), which comes from the Hall corpus (Hall and Tirre, 1979), the domain of quantification is understood to be restricted just to the children in the relevant class, not each child in existence.

(171) Each child was asked to bring in different kind of bread and she kept raving

about french (4;09)

It would be somewhat redundant to highlight this restricted domain with a relative clause. So the reason parents rarely linguistically restrict the domain in utterances with *each NP* may be due to the fact that *each* is used to quantify over local domains that are already contextually restricted.

4.1.5   Tense

The tense of the clause in which the quantifier appears may also be relevant. As seen in Table 4.5, most utterances are in present tense. But the rate of utterances in present tense is higher when universal quantification is indicated by *every* than when it is indicated by *each* ($\chi^2 = 17.2, p < .005$). In *each*-utterances, the quantifier is more likely to appear in a clause in past tense ($\chi^2 = 10.1, p < .01$) or in a "tenseless" clause ($\chi^2 = 54.1, p < .001$), a category that includes future-oriented imperatives like *put sugar in each coffee.*

Det	Total	Present	Past	Tenseless	Future	N/A
each	542	291	**103**	**117**	27	4
every	768	**556**	93	52	19	48

Table 4.5: Tense of the phrase in which the quantifier phrase appears in child-ambient speech. Particularly relevant values are in bold.

The prevalence of *every* in present tense is relevant to the genericity asymmetry in the following sense: Past tense is generally not used to express generic generalizations (perhaps due to the low likelihood of observing patterns that no longer hold

in the present).[4] For example, while (172a) is naturally understood to be a claim about dogs in general, (172b) is not. Instead, it is more naturally understood as a claim about some particular dogs in a contextually restricted domain.

(172)   a. Every dog barks

          b. Every dog barked

The same can be said of "tenseless" imperatives. Requests like those in (173), which all come from the Sachs corpus (Sachs and Nelson, 1983), are instructions to the child to change something about a very local domain, not to project beyond it.

(173)   a. Put a flower on each of the plates   (1;10)

          b. Put sugar in each coffee   (1;10)

          c. Put one leg in each side   (1;11)

So if parents tend to use *every* for expressing generalizations that project and tend to use *each* for expressing local generalizations, it makes sense that *every* most often occurs in present tense and that *each* is used relatively more frequently for reporting on past events and making requests.

4.1.6   Syntactic position of the quantifier phrase

Finally, the syntactic position of the quantifier phrase was coded. As seen in Table 4.6, *each* most often appears as part of an argument of a verb (e.g., *each*

---

[4] To be sure, past tense can be used with generic interpretations. Since dinosaurs are extinct, *dinosaurs laid eggs* can be understood as a statement about dinosaurs in general. Though as discussed in section 3.4, these sorts of kind attributions are not the sort of generics at issue here. Examples more relevant for the current topic include *Spartans valorized warriors* and *every Spartan was trained in battle*, both of which are naturally understood as generic in the relevant sense. In any case, these sorts of generalizations are not the norm in child-ambient speech.

toy needs to be put away or *put away each of your toys*). In contrast, *every* most often appears as part of an adjunct (e.g., *every time you play, you leave a mess*). Statistically, *each* is more likely than *every* to appear as part of a verb's argument ($\chi^2 = 182.0, p < .001$) and *every* is more likely than *each* to appear within an adjunct ($\chi^2 = 253.1, p < .001$).

Det	Total	Argument	Adjunct	N/A
each	542	**488**	36	18
every	768	254	**482**	32

Table 4.6: Syntactic position of the quantifier phrase in child-ambient speech. Particularly relevant values are in bold.

This effect is in large part driven by parents' propensity to use *every* to quantify over times, in which case *every NP* surfaces as a temporal adjunct. But the asymmetry between *each* and *every* persists even when all cases of quantification over times are excluded from analysis. That is, *every* is still more likely than *each* to appear inside of an adjunct ($\chi^2 = 39.7, p < .001$; Table 4.7).

QP	Total	Argument	Adjunct	N/A
each NP$_{\neg time}$	510	**483**	10	17
every NP$_{\neg time}$	282	227	**38**	17

Table 4.7: Excluding cases of quantifying over times, syntactic position of the quantifier phrase in child-ambient speech. Particularly relevant values are in bold.

This relative likelihood of using *every NP* as an adjunct may reflect parents using it to highlight projectable generalizations. Consider, for example, (174), from the NewmanRatner corpus (Newman et al., 2016), where the addition of the adjunct containing *every other baby* introduces a generalization that projects beyond the

220

local domain.

(174) Mother: Look here's a spoon.

M: What are we going to put on the spoon?

Child (0;11): [vocalizes]

M: You don't know but you'll stick it in your mouth just like every other baby

In contrast, if *each* is most often used to quantify over and predicate something of individuals in a very local domain, it makes sense that *each NP* should most often appear as an argument of a verb.

4.2   Sketching a learning story

To summarize the preceding section: There are low-level differences between *each* and *every* that are present in child-ambient speech. These include differences in what gets quantified over, how often the domain is explicitly restricted with a relative clause, tense, and syntactic position. These concomitants plausibly stem from parents using *each* to express generalizations over things in a local domain and using *every* to express generalizations that project beyond it. And if learners are sensitive to these differences, they might be able to use them to infer the correct meanings of *each* and *every*. This section sketches a proposal for how. It involves the elements listed in (175).

(175)   a. **Meaning representations**

each: First-order universal concept   (i.e., $\forall x(\mathbf{I}x)[\mathbf{E}x]$)

every: Second-order universal concept   (i.e., $\iota G(\forall x[Gx \equiv \mathbf{I}x])[\mathbf{E}(G)]$)

b. **Supporting cognitive systems**

each: Object-file representations

every: Ensemble representations

c. **Parents' intended message**

each: Generalize over local domain

every: Project beyond local domain

d. **Surface-level differences**

each: e.g., Quantify over individuals

every: e.g., Quantify over times

In a nutshell, the idea is as follows. From the parents' point of view, the goal in uttering a universally quantified statement is to convey some message (i.e., the speaker meaning, (175c)). Because ensembles better support projecting beyond the local domain – as discussed at the end of Chapter 3 – and because *every*'s meaning is an instruction to create an ensemble representation, they choose *every* if the intended message is a generic one. The low level differences, (175d), arise as a result of the nature of the intended message (e.g., situations are more likely to be generalized over in a way that projects; individuals are more likely to be generalized over in a way that remains local).

Upon hearing a universally quantified sentence, learners' task is to infer both their parents' intended message (i.e., the speaker meaning, (175c)) and the semantic representation of the quantifier used (i.e., the relevant universal concept, (175a)). Regarding the first goal, learners make inference from (175d) to (175c), which causes

them to represent the domain of quantification using one of the two cognitive systems in (175b). Regarding the second goal, they notice that utterances of *each* correlate with representing the domain as object-files whereas utterances of *every* correlate with representing the domain as an ensemble. This difference leads them to pair *each* with a pre-existing first-order universal concept and *every* with a pre-existing second-order universal concept.

At this point, it is worth noting that this sort of learning story (and cognitivism about meanings in general) requires a certain conception of the semantics/pragmatics distinction that is not universally adopted. On cognitive views of meaning, sentences don't determine (or have) truth-conditions, they just provide some constraints on thought-building. As Carston (2008) puts it, the meaning, at best, often offers a "gappy" proposition or a "schema or template for building propositions" (p.324). Pragmatics – in a broad sense that includes theory of mind, coordination of social behavior, and world knowledge – must work alongside meaning to "fill in the gaps." The end result is a thought. Its contours are shaped both by the semantic representation and the pragmatics.[5]

With that in mind, we can consider the links in the proposed chain of inference: (175d) → (175c), (175c) → (175b), and (175b) → (175a). Each link requires elaboration, to which we now turn. Each link also requires empirical defense, about which some suggestions are offered in section 4.3.

Regarding the first link, this chapter discussed reasons for thinking the ob-

---

[5] Compare this picture to one in which semantics yields a full truth-evaluable proposition and the role of pragmatics is to enable further inferences about the speakers' mental state, given that they conveyed that proposition and not some other proposition.

served surface-level differences are related to a difference in intended speaker meaning. Learners may reason along the same lines. If a learner hears their parents quantify over times, for example, they may be biased toward a generic understanding of the utterance more than they would be if they hear their parents quantify over locally-present objects. Here, work arguing that preschoolers "default" to generic thinking (Gelman 2009; Leslie and Gelman 2012; Leslie 2012) may be relevant. If right, this work suggests that learners' null hypothesis is one compatible with projection. So children may use cues like quantification over times as a confirmation of an initial guess that their parent intended to project beyond the local domain. Likewise, quantification over a small number of locally-present individuals may serve to disconfirm that pre-potent hypothesis about their parents' intended message.

In forming a guess about their parents' intended message, learners will rely either on object-file or ensemble representations.[6] As discussed in section 3.4, ensemble representations allow for projecting beyond the local domain whereas object-file representations do not. Namely, ensembles are initiated based on homogeneity and to be a member of an ensemble is to contribute to its summary statistics. This naturally licenses predictions about potential new members: They will contribute in the same way (e.g., be a similar color, be a similar size, be a similar orientation,

---

[6] Given the frequency of quantification over times with *every*, this proposal assumes that ensemble representations are formed when representing large collections of events. Event perception and ensemble representation have not yet been studied in tandem. But given the wide range of things that routinely are represented as ensembles – objects, faces, sounds – there is reason to think the same format of representation may apply to events as well. The idea of representing events as ensembles accords with intuition as well. Try to recall every time you went grocery shopping. You likely don't recall individual episodic memories (unless something exceptional happened at the grocery store). Instead, you likely remember summary statistics: the average duration of the trips, your average average mood while shopping, etc.

etc.). Object-files, in contrast, are pointers to individuals. Two object-files are two distinct pointers, and that the objects they point to share any properties is an accident from this system's point of view.

So representing an intended message – i.e., building a thought – that suggests projection beyond the local domain will implicate the system of ensemble representation. On the other hand, object-file representations are well-suited for use in representing an intended message in which something is predicated of individuals in a local domain. Shared-attention to what is being quantified over may play a role in triggering the relevant representational system perceptually. For instance, if a parent says *put a napkin next to each plate* in the presence of a dinner table, the child may represent the plates as object-files, which may itself support a local understanding of the request in addition to providing an excellent learning opportunity.

It seems less likely that perceptual information of this sort will trigger ensemble representations when a parent says e.g., *every time we eat, we use napkins*. But perceptual triggering of ensemble representations may happen in cases where *every* is used to quantify over individuals in a way that is meant to project (e.g., if the parent said *every picture you drew this week was that color* in the presence of a dozen finger paintings). In the absence of such perceptual cases, one possibility is that ensemble representations are required for thinking generic thoughts. This constitutes a substantive hypothesis about the relationship between ensembles and projection, which has not yet been explored (see note 6).

The final link – between the supporting cognitive systems and the meaning

representations – was the main contention of Chapter 3. In a sense, this is the easiest link from the learner's perspective, as it is baked into the proposed first-order and second-order representations that serve as the meanings of *each* and *every*. The first-order universal concept is an individual-based thought, and consequently, it naturally interfaces with object-file representations. The second-order universal concept is a group-based thought, and consequently, it naturally interfaces with ensemble representations. The present account assumes that both of these concepts are available pre-linguistically (and in another line of work, we have begun to investigate how early these concepts are available to infants (Cesana-Arlotti et al., 2020)).

But even assuming these two concepts are in the learner's initial hypothesis space, more needs to be said about why those concepts are the main hypotheses under consideration. Learners likely narrow their hypothesis space initially through Syntactic Bootstrapping (Gleitman, 1990). In particular, distributional syntactic information can help them infer that *each* and *every* are determiners that express quantity-information as opposed to, say, adjectives that specify property-information (Syrett et al. 2012; Wellwood et al. 2016).

Perhaps, in line with Chapter 2, the hypothesis space is further limited by a universal constraint on possible determiner meanings: They are restricted. This would exclude some possible second-order hypotheses that might otherwise be considered for *every* (e.g., "$\subseteq$"). (In fact, Chapter 2 predicts that no determiners express relations, which constrains learners' hypothesis space not just for the universals but for quantificational determiners in general.)

In addition, learners will need some way of ruling out possible determiner meanings that make the wrong truth-conditional contribution (e.g., an existential or proportional representation). This may require coordination of linguistic evidence with the co-present referent world, which has been shown to play an important role in acquiring other abstract words, including mental state verbs (Papafragou et al., 2007) and negation (Gomes et al., 2020).

The above proposal has been concerned with the question of how learners decide between a restricted first-order universal concept and a restricted second-order universal concept when acquiring *each* and *every*. As noted, many aspects of this proposal await empirical support. This highlights a number of potential research avenues, some of which are discussed in the following section.

4.3   Future directions: Testing the proposal

One obvious question raised by this proposal is whether learners actually are sensitive to these particular linguistic cues. And if so, which ones are particularly important for learning? It might be, for example, that the times vs. individual distinction serves as a strong signal to learners, but that the difference in tense does not. Alternatively, it might be that the past tense or future-oriented imperatives with *each* are crucially important. If learners do start with a bias to expect that their parents will project – in line with the "generics as default" hypothesis noted above (Gelman 2009; Leslie and Gelman 2012; Leslie 2012) – then cases of quantification over individuals might serve as a sort of "eureka moment" for learners,

forcing them to abandon their initial hypothesis and consider the possibility that the generalization was intended to be locally bounded.

In this context, the role for non-linguistic cues may be especially important. As discussed in section 3.2 of the previous chapter, object-files are subject to working memory limits whereas ensembles are not.[7] And if a small number of objects are present, object-file representations are likely to be perceptually triggered. So instances of parents quantifying over small numbers of locally present objects might serve as "perceptual gems" (Carey and Bartlett 1978; Heibeck and Markman 1987; Medina et al. 2011) that push learners to consider the first-order universal concept as a hypothesis about *each*'s meaning.

These possibilities all correspond to distinct hypotheses about (175d), the surface-level differences that arise from the different uses of *each* and *every*. One potential route for exploring them further would be a novel quantifier learning experiment with children (Hunter and Lidz 2013; Wellwood et al. 2016). One group of participants might be asked to learn a nonce word – *gleeb* – presented an *each* context, like (176a), while another group is asked to learn the same word from an *every* context, like (176b).

(176)   a. The teacher gave gleeb one of the kids a snack   (*each*)

        b. Gleeb time they eat, the teacher gives the kids a snack   (*every*)

Assuming children successfully learn the truth-conditions, we can ask whether they

---

[7] That is, while the number of ensembles that can be represented simultaneously is subject to working memory limitations (Halberda et al., 2006), the number of objects that constitute a single ensemble is not.

paired *gleeb* with a first-order or a second-order concept by following-up this experiment with a *gleeb* version of the center-of-mass experiment from section 3.3.3 of Chapter 3. The exact details of the context could then be varied – e.g., including or excluding relative clauses, tense differences, or quantification over times – to determine which combination of cues best supports learning.

Similar conclusions about which cues matter in principle could be drawn from a series of Human Simulation Paradigm experiments on adults (Gillette et al., 1999).[8] In these tasks, adult participants view short vignettes of parents uttering a sentence (containing *each* or *every*, in this case), with the video muted. A tone is heard during the duration of the sentence, and the participant is asked to guess what they think the parent said during that beep. The rate at which participants type sentences containing *each* or *every* can be taken as an estimate of the extent to which an observer, armed only with the referent world and the time of the utterance, could ever infer the presence of quantification in the message.

The usefulness of various linguistic cues could then be estimated by including a version of the sentence with the critical quantifier removed. To simulate younger stages of development, the content words could also be replaced by meaningless novel words. In a scene where the parent said (177a), for example, participants might be presented with something like (177b).

(177)   a. Pick up each toy in your room!

         b. Meech up ＿＿＿ zeb in lo fendle!

---

[8] This possibility is currently being explored in collaboration with Victor Gomes.

This method allows for controlling which aspects of the linguistic stimulus participants have access to. For example, one version might grant participants access to information about what is being quantified over by replacing *zeb* with *toy*. To the extent that a cue is in principle helpful for learners, its inclusion will be helpful for adults trying to guess the message.

Of course, the low-level differences that ultimately matter in English may not be the same differences that matter cross-linguistically. That is, distributional footprints of projecting beyond the local domain may differ language-to-language. What is predicted by the proposed account is just that if a language has a second-order universal quantifier, it should be compatible with the relevant class of generic generalizations. Likewise, if a language has a first-order universal quantifier, it should not support projecting beyond the local domain. But how exactly these projectable or local thoughts routinely get expressed may be subject to variation.

This raises questions about the relationship between the surface-level differences and parents' intended message (the (175d) $\rightarrow$ (175c) link). Initially, at least, one way forward on this front would be to conduct corpus analyses of child-ambient speech in other languages. Findings could be paired with experiments like those reported in Chapter 3, to support particular hypotheses about different determiners' mental representations (for example, Spanish *cada* is often translated as *each*, but if it systematically causes ensemble representation, it may be better thought of as analogous to *every*, which would lead to different predictions about a corpus analysis of Spanish).

Moving up the chain, future work might also aim to provide support for the

230

proposed relationship between the intent to project beyond the local domain and the supporting cognitive systems (the (175c) → (175b) link). Section 3.4 in the previous chapter hinted at some predictions pertaining to this connection. For example, if participants are encouraged to represent a group of objects as an ensemble – e.g., a group of green circles – they should be more likely to project beyond the initial domain and make a prediction about the next object they will encounter (e.g., that it will be green). If they are instead encouraged to treat the same group as a series of object-files, they should be less willing to make the same prediction. Though controlling for adults' ability to recognize patterns in this sort of task may be difficult, a result along these lines would provide strong evidence for this particular component of the acquisition proposal.

Lastly, the focus on the genericity asymmetry may offer a new window onto the age of acquisition debate mentioned at the beginning of this chapter. In particular, we might ask: How early are learners sensitive to the fact that *every* but not *each* supports projectable generalizations? To address this, children might be shown novel objects labeled with a novel noun (e.g., *dax*) that is either quantified over by *each* or *every*, as in (178).

(178)   a. Each dax has green hair

          b. Every dax has green hair

Children who hear something like (178b) are predicted to generalize that the next dax they encounter will also have green hair.

This propensity to generalize might be pitted against children's strong bias

231

to generalize novel labels on the basis of shape (Landau et al., 1998, 1992, 1988). For example, imagine that the original daxes are circles with green hair, and that a child is presented with a square with green hair and a circle with no hair and asked "Which one is also a dax?" They may be more inclined to violate the shape bias (and pick the green-haired square) if the daxes were introduced with (178b) than if they were introduced with (178a). Figuring out when children show this effect could help fix an upper bound on the age of acquisition.[9]

In sum, the learning story sketched in section 4.2 relies on an idea that is not common in work on language acquisition: using the extralinguistic systems (object-files and ensembles) connected with sentences as a source of evidence for inferring meaning. This proposal makes a range of rich predictions that will be (or are currently being) explored. So while it is too early to say the proposed first-order and second-order representations for *each* and *every* are supported by a fully-fledged account of their acquisition, it at least seems fair to say that thinking about their meanings at this level of detail reveals a vista of new empirical directions.

4.4 Chapter summary

This chapter considered the acquisition of *each* and *every*. Given the representations proposed in the preceding chapters, learners' task in acquiring these terms is to pair one pronunciation with one universal concept and another pronunciation with a logically-equivalent but representationally distinct universal concept. This

---

[9] This possibility is currently being explored in collaboration with John Trueswell and Anna Papafragou

places a larger burden on the child and on the theorist than a view on which the target of acquisition is the truth-conditional content of the determiner, represented in no particular format.

Of course, determining how children arrive at universal truth-conditions is a difficult question in its own right. In fact, determining when children correctly represent these truth-conditions is difficult enough, and is the subject of much debate. Abstracting away from both of these questions, this chapter asked what data might be available in speech to children that could signal the relevant difference between *each* and *every* and how that data might be used by learners.

Section 4.1 reported results of a corpus analysis suggesting that the genericity asymmetry discussed in Chapter 3 is particularly relevant for language acquisition. Parents seem to use *each* to express local generalizations (e.g., *you've got a glove on each hand*) and *every* to make generalizations that project beyond the local domain (e.g., *every time we go outside, you forget your gloves*). This difference in intended message leaves low-level distributional footprints (e.g., in terms of what gets quantified over and whether the quantifier phrase appears within an adjunct or as a verb's argument).

Section 4.2 then sketched a proposal for how learners might use these observable differences to ultimately infer the right meanings of *each* and *every*. The chain of inference starts with low-level differences and transits through parents' intended messages and the cognitive systems that support them, which are linked to the proposed first-order and second-order meaning representations. The proposal assumes that learners have access to the relevant concepts pre-linguistically and that their

hypothesis space is also narrowed in more traditional ways (e.g., thanks to syntactic bootstrapping and universal grammar).

At the moment, the proposal outlined above is just that. But as section 4.3 laid out, thinking about the mental representations of *each* and *every* in this way opens the door for a variety of new empirical explorations. To the extent that they produce the predicted results, the proposed representations for *each* and *every* will be supported. And at the least, focusing on representational format has allowed for thinking about the acquisition question in a new light.

In particular, the question of how sentences relate to extralinguistic cognition has not generally been taken to be relevant to language acquisition. It is often assumed that a large aspect of learning meanings is tracking situational context: that parents say *dog* in the presence of dogs and *frog* in the presence of frogs and that learners track these co-occurrences (Smith and Yu, 2008).[10] Since Gleitman (1990), the importance of syntactic cues, working in concert with the situational context, has also been clear (e.g., Naigles 1990; Lidz et al. 2003; Yuan and Fisher 2009). Moreover, Hacquard and Lidz (2018) raise the possibility that pragmatic factors – in particular connections between sentence meaning and speaker meaning – can supplement syntactic bootstrapping. And in a similar vein, Dudley (2017) discusses the role of speakers' conversational goals in the inferences learners draw about meanings.

The above proposal can be thought of as adding a new dimension to this list:

---

[10] With quantifiers the situation is more complex, but nonetheless, the focus has been on learning the right truth-conditions and, for example, distinguishing universals from existentials.

234

details about how sentences are related to the cognitive systems with which they interface. This type of information can be considered when other avenues (i.e., the type of information mentioned above) are insufficient to lead learners to the target meaning. And importantly, this addition to the acquisitionist's arsenal is a direct result of treating the target of learning as not just truth-conditional content, but content represented in a particular format.

Chapter 5: Conclusion

This dissertation has been concerned with the mental representation of the universal quantifiers *each* and *every* (and, to a lesser extent, *all*). Their truth-conditional contribution can be specified by theorists in many ways. Chapter 1 argued that these logically equivalent formalisms should be treated as psychologically distinct hypotheses.

For example, on a psychological interpretation of Generalized Quantifier Theory (Barwise and Cooper, 1981), quantificational determiners might be said to express relations between groups. On this view, the mental representation of *every frog is green* would be something like *the frogs are among the green things*. In contrast, the same sentence might be represented in a non-relational way that doesn't implicate the green things as an independent group, as in *the frogs are such that they are green*. Moreover, it might be represented in a non-relational way that doesn't implicate the frogs or the green things as a group, as in *any individual that's a frog is such that it is green*.

Chapter 2 argued that quantificational determiners are better thought of in non-relational terms. The psychosemantic evidence marshaled in support of this claim was that in evaluating sentences like *every big circle is blue*, participants only

236

represent the big circles as a group. Neither the blue circles nor the big blue circles are treated in the same way. This result is not predicted if *every* is represented as a device for relating two groups. If both groups are logically on a par, they should be treated psychologically on a par as well. But the result is predicted if *every* is represented as a device for combining with an internal argument to form a restricted quantifier, as this view makes it clear that the two arguments – *big circle* and *is blue* – play distinct logical roles.

Semantically, this non-relational alternative was argued to be preferable because it allows for a simple explanation of the "conservativity" constraint (Barwise and Cooper 1981; Higginbotham and May 1981; Keenan and Stavi 1986). This cross-linguistic universal states that all determiners permit duplicating their internal argument in their external argument without a change in truth-conditions. For example, *every frog is green* iff *every frog is a frog that is green*. Given that there are plenty of relations that disobey this constraint, this is a puzzling generalization if determiners express relations. But it is a logical consequence if *every frog* and the like express restricted quantifiers (Pietroski 2005, 2018; Westerståhl 2019; Lasersohn 2021).

Chapter 3 argued that while *every* and *all* have representations that implicate a group, *each* has a representation that is completely first-order in that it implicates only individuals. A parallel was drawn between first-order and second-order quantification in language – a distinction in what types of variable positions are quantified into – and individuals and groups in non-linguistic cognition. In particular, first-order quantification was proposed to serve as an instruction to cognition to

represent the domain as independent object-files whereas second-order quantification was proposed to serve as an instruction to cognition to represent the domain as an ensemble. This connection was leveraged in a series of experiments showing that when evaluating sentences like *every circle is green*, adults and children represent the circles as an ensemble. But when evaluating *each circle is green* with respect to the same image, adults and children instead treat the circles as independent object-files.

This first-order/second-order distinction was argued to explain a genericity asymmetry between *each* and *every* (Beghelli and Stowell, 1997). In particular, *every frog is green* can be understood as a claim about frogs in general, whereas *each frog is green* seems to be unable to project beyond the local domain. Instead of situating this distinction in grammar, it was argued that this distinction is a result of the connection between *each*'s first-order representation and object-files and *every*'s second-order representation and ensembles. Properties of these systems make ensembles more conducive to projecting (i.e., licensing predictions) beyond the local domain.

Chapter 4 built on the proposed difference between *each* and *every* and asked how they might be acquired by children. Investigating a corpus of child-ambient speech revealed that parents use *each* to state generalizations over local domains (e.g., *make sure you have shoes on each foot*) and use *every* to state broader generalizations that project beyond it (e.g., *every time we get your shoes on, you cry*). These differences in intended message leave low-level footprints. For example, *each* is far more likely to be used to quantify over individuals, whereas *every* is more likely to be used to quantify over times.

Given this data, an acquisition story was sketched. It relied on learners using details of extralinguistic cognition – object-files and ensembles – to help make the inference from the observed surface-level differences to the right universal concepts to pair with *each* and *every*. This represents a new source of evidence that learners might use, one that is not usually discussed in research on language acquisition. The proposal also suggested a range of follow-up experiments, aimed at testing various novel predictions.

5.1   Methodological implications

Finding evidence for one particular representation over another is difficult. Claims of semantic decomposition are not only hard to support experimentally, they are also are hard to shield from alternative explanations. The preceding chapters highlighted methodological strategies for dealing with both of these issues.

In terms of supporting decomposition experimentally, the right linking hypothesis is crucial. All of the experiments reported above rely on – and provide further support for – the Interface Transparency Thesis (Lidz et al., 2011). This is the idea that details of a sentence's semantic representation carry detectable weight in determining which strategies competent speakers will use to evaluate that sentence. For this linking hypothesis to work as a methodological strategy, the interfacing systems need to be well-understood. The fact that notions like object-files, ensembles, and cardinality are so well-established has made quantificational determiners an especially fertile testing ground for this linking hypothesis.

In terms of fending off alternative explanations, the first task is meeting Fodor's challenge: A retreat to strongly associated world knowledge is always available (Fodor et al. 1975; Fodor and Lepore 1998; Fodor 1998). For example, an experiment finding that participants represent ANIMAL any time they represent *frog* would not be compelling evidence that the meaning of *frog* in part decomposes into ANIMAL. The lexical item *frog* might just have the atomic concept FROG as its meaning, with the fact that frogs are animals being a bit of associated world knowledge. Why think the situation is any different for *frog* and ANIMAL or *doe* and DEER than it is for the hypothetical finding that participants routinely represent PEPPER upon hearing *salt*? It's plausible that the two differ only in our understanding of the relation between the extensions of the two concepts: necessary inclusion in the first case, frequent co-occurence in the second.

Cases of "logical" vocabulary, like *every* and *most*, are able to avoid this problem because the proposed decompositions implicate concepts that we would not otherwise expect to be strongly associated with the lexical item. A retreat to associated world knowledge is still available, but "restricted second-order quantification" is at least a less likely candidate for a strong association than something like "does are deer." After all, you might encounter "does are deer" when you learn the word *doe*, or when you look it up in a dictionary, or when you sing the song. But the same is clearly not true for *each* and *every*.

In sum, good candidates for semantic decomposition will meet at least the following two criteria. First, the proposed decomposition will interface with cognitive systems whose properties are well-understood. And second, the proposed

decomposition will be in terms of concepts that are not be predicted by association alone.

5.2   Theoretical implications

Language connects pronunciations and meanings. From the outset, this dissertation adopted a mentalistic view of what pronunciations and meanings are: Pronunciations relate to motor-planning systems and articulators and meanings relate to non-linguistic cognitive systems and concepts (Chomsky, 1964). In particular, a meaning representation provides instructions to cognition for assembling a thought (Pietroski, 2018).

Chapter 1 raised some initial questions that arise when thinking about meanings in this mentalistic way: What sorts of instructions do meanings provide to cognition? At what grain-size are these instructions shared by competent speakers? And to what extent do the instructions supplied by the meaning constrain the contours of the resulting thought?

We can imagine a number of possible answers to this family of questions, even restricting our attention to the "logical" vocabulary. One possibility is that speakers share an understanding of the informational contribution that lexical items make, but they represent this contribution in various ways depending on the situation. In other words, a lexical item like *every* is represented as an equivalence class of truth-conditionally equivalent specifications, so its meaning restricts thought-building only insofar as it restricts truth-conditions.

Another possibility is that speakers each represent the informational contribution of lexical items in one particular format, but individuals differ with respect to which format. Whereas you might represent *every* in restricted terms, maybe I represent it in relational terms. The two of us never disagree in conversation, but our mental lives differ. In other words, the meaning of a lexical item like *every* offers a precise but idiosyncratic instruction to cognition, which differs for each speaker.

A third possibility is that speakers represent the informational contribution in a shared representational format, but that format is unstructured. For example, *every* just means EVERY and so it's no surprise that everyone represents it in the same way. In this case, the meaning representation provides an extremely constraining instruction to cognition, but a relatively uninteresting one.

Given any of these answers, we might be tempted to abstract away from the psychological details of meaning representations, or to treat these details as secondary phenomena when constructing theories of meaning.

But the results presented here paint a different picture. What serves as the meaning of *every* is a representation with a very particular format. Competent speakers seem to share this representation. This is surprising because, in principle, two speakers could have acquired a different representation that made the same informational contribution, and they would never disagree in conversation. Perhaps even more surprising is that the apparently shared meaning representation has internal structure and is not (an instruction to access) an atomic concept. So in this case at least, the meaning seems to provide a precise instruction to cognition (e.g., about which and how many groups to represent) that is shared by speakers and that

has discoverable internal structure.

This suggests that representational format is not a secondary phenomenon, but one that is crucial to consider in giving a full account of linguistic meaning.

Appendix A: Corpora used in Chapter 4

The following corpora from the North American English portion of CHILDES were included in the analysis reported in Chapter 4. For each corpus, the number of children, age range, and citation (when applicable) is included.

- Bates: 27 children (1;8 & 2;4) Bates et al. (1991)
- Bernstein Ratner: 9 children (1;1 - 1;11) Bernstein (1982)
- Bliss: 7 children (2;3, 2;5, 3;4, 4;3, 4;6, 5;4, 6;1) Bliss (1988)
- Bloom 1970: 3 children (1;1 - 3;2) Bloom (1970); Bloom et al. (1974, 1975)
- Bloom 1973: 1 child (1;4 - 2;10) Bloom (2013)
- Bohannon: 2 children (2;8 & 3;0) Bohannon III and Marquis (1977)
- Braunwald: 1 child (1;0 - 6;0) Braunwald (1971)
- Brent-Siskind: 16 children (0;6 - 1;0) Brent and Siskind (2001)
- Brown: 3 children (1;6 - 5;1) Brown (1973)
- Clark: 1 child (2;2 - 3;2) Clark (1978)
- Cornell: 8 children (1;6 - 4;0)
- Demetras-Trevor: 1 child (2;0 - 3;11) Demetras (1989b)
- Demetras-Working: 3 children (2;0 - 2;6) Demetras (1989a)
- EllisWeismer: 138 children (2;6, 3;6, 4;6, & 5;6) Heilmann et al. (2005)
- Evans: 44 children (2;1 - 2;9)
- Feldman: 1 child (0;5 - 2;9) Feldman (1998)
- Garvey: 48 children (2;10 - 5;7) Garvey and Hogan (1973)

- Gathercole: 8 children (2;9 - 6;6) Gathercole (1986)

- Gelman: 118 children (1;6 - 7;0) Gelman et al. (1998, 2004)

- Gleason: 24 children (2;1 - 5;2) Gleason (1980); Masur and Gleason (1980)

- Haggerty: 1 child (2;6) Haggerty (1930)

- Hall: 39 children (4;6 - 5;0) Hall and Tirre (1979)

- Higginson: 3 children (0;11 - 2;11) Higginson (1985)

- HSLLD: 83 children (2;0 - 6;0) Dickinson and Tabors (2001)

- Kuczaj: 1 child (2;4 - 4;1) Kuczaj II (1977)

- MacWhinney: 2 children (0;7 - 8;0) MacWhinney (1991)

- McCune: 9 children (1;0 - 3;0) McCune (1995)

- McMillan: 2 children (3;0)

- Morisset: 205 children (2;6 - 3;6) Morisset et al. (1990)

- Nelson: 1 child (1;9 - 3;0) Nelson (1989)

- New England: 52 children (1;2, 1;8, & 2;3) Ninio et al. (1994)

- NewmanRatner: 121 children (0;7 - 2;0) Newman et al. (2016)

- Nicholas: Normal Hearing: 45 children (1;0 - 4;0) Nicholas and Geers (1997)

- Peters/Wilson: 1 child (1;7 - 4;1) Wilson and Peters (1988)

- POLER (control): 26 children (5;0 - 7;0) Berl et al. (2005); Gaillard et al. (2007); Mbwana et al. (2009)

- Post: 3 children (1;7 - 2;8) Demetras et al. (1986)

- Rollins: 12 children (0;6 - 1;0) Rollins and Trautman (2011); Rollins (2003); Rollins and Greenwald (2013); Trautman and Rollins (2006)

- Rondal: Normal: 21 children (1;8 - 2;8) Rondal (1976)

- Sachs: 1 child (1;1 - 5;1) Sachs and Nelson (1983)

- Sawyer: 20 children (3;6 - 4;11) Sawyer (2013)

- Snow: 1 child (2;5 - 3;9) MacWhinney and Snow (1990)

- Soderstrom: 2 children (0;6 - 0;10) Soderstrom et al. (2008)

- Suppes: 1 child (1;11 - 3;3) Suppes (1974)

- Tardiff: 25 children (1;6 - 1;9)

- Valian: 21 children (1;9 - 2;8) Valian (1991)

- VanHouten: 28 children (2;0 & 3;0) Van Houten (1986)

- Van Kleeck: 20 children (3;0 - 4;0)

- Warren: 20 children (1;6 - 6;2) Warren-Leubecker (1982)

- Weist: 6 children (2;1 - 5;0) Weist and Zevenbergen (2008)

Bibliography

Achimova, A., Syrett, K., Musolino, J., and Déprez, V. (2017). Childrens developing knowledge of wh-/quantifier question-answer relations. *Language Learning and Development*, 13(1):80–99.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of Hirotugu Akaike*, pages 199–213. Springer.

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

Alvarez, G. A. and Franconeri, S. L. (2007). How many objects can you track?: Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision*, 7(13):14–14.

Alvarez, G. A. and 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.

Bae, G.-Y., Olkkonen, M., Allred, S. R., and 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.

Barwise, J. and Cooper, R. (1981). Generalized quantifiers and natural language. In *Philosophy, Language, and Artificial Intelligence*, pages 241–301. Springer.

Bates, E., Bretherton, I., and Snyder, L. S. (1991). *From First Words to Grammar: Individual Differences and Dissociable Mechanisms*, volume 20. Cambridge University Press.

Beghelli, F. (1997). The syntax of distributivity and pair-list readings. In Szabolcsi, A., editor, *Ways of Scope Taking*, pages 349–408. Springer.

Beghelli, F. and Stowell, T. (1997). Distributivity and negation: The syntax of each and every. In Szabolcsi, A., editor, *Ways of Scope Taking*, pages 71–107. Springer.

Benmamoun, E. (1999). The syntax of quantifiers and quantifier float. *Linguistic Inquiry*, 30(4):621–642.

Berl, M., Balsamo, L., Xu, B., Moore, E., Weinstein, S., Conry, J., Pearl, P., Sachs, B., Grandin, C., Frattali, C., et al. (2005). Seizure focus affects regional language networks assessed by fmri. *Neurology*, 65(10):1604–1611.

Bernstein, N. E. (1982). *Acoustic Study of Mothers Speech to Language-learning Children: An Analysis of Vowel Articulatory Characteristics*. Boston University dissertation.

Bliss, L. S. (1988). Modal usage by preschool children. *Journal of Applied Developmental Psychology*, 9(3):253–261.

Bloom, L. (1970). *Language Development: Form and Function in Emerging Grammars*. MIT Press.

Bloom, L. (2013). *One Word at a Time: The Use of Single Word Utterances Before Syntax*, volume 154. Walter de Gruyter.

Bloom, L., Hood, L., and Lightbown, P. (1974). Imitation in language development: If, when, and why. *Cognitive Psychology*, 6(3):380–420.

Bloom, L., Lightbown, P., Hood, L., Bowerman, M., Maratsos, M., and Maratsos, M. P. (1975). Structure and variation in child language. *Monographs of the Society for Research in Child Development*, pages 1–97.

Bobaljik, J. D. (2012). *Universals in Comparative Morphology: Suppletion, Superlatives, and the Structure of Words*. MIT Press.

Bohannon III, J. N. and Marquis, A. L. (1977). Children's control of adult speech. *Child Development*, pages 1002–1008.

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.

Braunwald, S. R. (1971). Mother-child communication: The function of maternal-language input. *Word*, 27(1-3):28–50.

Brendel, C. I. (2019). An investigation of numeral quantifiers in english. *Glossa: A Journal of General Linguistics*, 4(1).

Brent, M. R. and Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, 81(2):B33–B44.

Brooks, P. J. and Braine, M. D. (1996). What do children know about the universal quantifiers all and each? *Cognition*, 60(3):235–268.

Brown, R. (1973). *A First Language: The Early Stages*. Harvard Univeristy Press.

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

Carey, S. (2009). *The Origin of Concepts*. Oxford Series in Cognitive Development. Oxford University Press.

Carey, S. and Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, 15:17–29.

Carston, R. (2008). Linguistic communication and the semantics/pragmatics distinction. *Synthese*, 165(3):321–345.

Cesana-Arlotti, N., Knowlton, T., Lidz, J., and Halberda, J. (2020). An investigation of the origin of logical quantification: Infant's and adult's representations of collective and distributive actions in complex visual scenes. *Poster presented at the 42nd Annual Virtual Meeting of the Cognitive Science Society.*

Champollion, L. (2017). *Parts of a Whole: Distributivity as a Bridge Between Aspect and Measurement*, volume 66. Oxford University Press.

Champollion, L. (2020). Distributivity, collectivity, and cumulativity. *The Wiley Blackwell Companion to Semantics*, pages 1–38.

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.

Cheries, E. W., Feigenson, L., Scholl, B. J., and Carey, S. (2005). Cues to object persistence in infancy: Tracking objects through occlusion vs. implosion. *Journal of Vision*, 5(8):352–352.

Chierchia, G. (1995). Lecture notes from a talk given at Utrecht university. *Unpublished.*

Chomsky, N. (1964). *Current Issues in Linguistic Theory*. De Gruyter Mouton.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. Greenwood Publishing Group.

Chong, S. C. and Treisman, A. (2003). Representation of statistical properties. *Vision Research*, 43(4):393–404.

Church, A. (1941). *The Calculi of Lambda Conversion*. Princeton University Press.

Clark, E. V. (1978). Awareness of language: Some evidence from what children say and do. In *The Childs Conception of Language*, pages 17–43. Springer.

Coppock, E. and Brochhagen, T. (2013a). Diagnosing truth, interactive sincerity, and depictive sincerity. In *Semantics and Linguistic Theory*, volume 23, pages 358–375.

Coppock, E. and Brochhagen, T. (2013b). Raising and resolving issues with scalar modifiers. *Semantics and Pragmatics*, 6:3–1.

Crain, S., Thornton, R., Boster, C., Conway, L., Lillo-Martin, D., and Woodams, E. (1996). Quantification without qualification. *Language Acquisition*, 5(2):83–153.

Davidson, D. (1967a). The logical form of action sentences. In Rescher, N., editor, *The Logic of Decision and Action*, pages 216–234. University of Pittsburgh Press.

Davidson, D. (1967b). Truth and meaning. In *Philosophy, Language, and Artificial Intelligence*, pages 93–111. Springer.

de Vries, H. (2015). *Shifting Sets, Hidden Atoms: The Semantics of Distributivity, Plurality and Animacy*. Utrecht University dissertation.

Dehaene, S. (2011). *The Number Sense: How the Mind Creates Mathematics*. Oxford University Press.

Demetras, M. (1989a). Changes in parents conversational responses: A function of grammatical development. *Paper presented at ASHA*.

Demetras, M. (1989b). Working parents' conversational responses to their two-year-old sons. *University of Arizona Working Papers*.

Demetras, M. J., Post, K. N., and Snow, C. E. (1986). Feedback to first language learners: The role of repetitions and clarification questions. *Journal of Child Language*, 13(2):275–292.

Demeyere, N., Rzeskiewicz, A., Humphreys, K. A., and Humphreys, G. W. (2008). Automatic statistical processing of visual properties in simultanagnosia. *Neuropsychologia*, 46(11):2861–2864.

Dickinson, D. K. and Tabors, P. O. (2001). *Beginning Literacy with Language: Young Children Learning at Home and School*. Paul H Brookes Publishing.

Donaldson, M. and Lloyd, P. (1974). *Sentences and Situations: Children's Judgments of Match and Mismatch*. Centre National de la Recherche Scientifique.

Dowty, D. (1987). Collective predicates, distributive predicates, and all. *Proceedings of the 3rd ESCOL*, pages 97–115.

Dowty, D. and Brodie, B. (1984). The semantics of "floated" quantifiers in a transformationless grammar. In *Proceedings of the West Coast Conference on Formal Linguistics*, volume 3, pages 75–90.

Dowty, D. R. (1979). *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ*. Springer.

Drozd, K. F. (2001). Children's weak interpretations of universally quantified questions. In Bowerman, M. and Levinson, S., editors, *Language Acquisition and Conceptual Development*, volume 3 of *Language Culture and Cognition*, pages 340–376. Cambridge University Press.

Dudley, R. E. (2017). *The Role of Input in Discovering Presupposition Triggers: Figuring Out What Everybody Already Knew*. University of Maryland dissertation.

Engelberg, S. (2011). Frameworks of lexical decomposition of verbs. In Meienborn, C., von Heusinger, K., and Portner, P., editors, *Semantics (HSK 33.1)*, pages 358–399. Walter de Gruyter.

Feigenson, L. and Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition*, 97(3):295–313.

Feigenson, L., Carey, S., and Hauser, M. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological Science*, 13(2):150–156.

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

Feiman, R. and Snedeker, J. (2016). The logic in language: How all quantifiers are alike, but each quantifier is different. *Cognitive Psychology*, 87:29–52.

Feldman, A. (1998). *Constructing Grammar: Fillers, Formulas, and Function*. University of Colorado at Boulder dissertation.

Fillmore, C. J. (1968). Lexical entries for verbs. *Foundations of Language*, 4:373–393.

Fodor, J. A. (1970). Three reasons for not deriving "kill" from "cause to die". *Linguistic Inquiry*, 1(4):429–438.

Fodor, J. A. (1983). *The Modularity of Mind*. MIT Press.

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

Fodor, J. A., Garrett, M. F., Walker, E. C., and Parkes, C. H. (1980). Against definitions. *Cognition*, 8(3):263–367.

Fodor, J. A. and Lepore, E. (1998). The emptiness of the lexicon: Reflections on james pustejovsky's the generative lexicon. *Linguistic Inquiry*, 29(2):269–288.

Fodor, J. D., Fodor, J. A., and Garrett, M. F. (1975). The psychological unreality of semantic representations. *Linguistic Inquiry*, 6(4):515–531.

Foster, J. A. (1976). Meaning and truth theory. In Evans, G. and McDowell, J., editors, *Truth and Meaning*. Oxford University Press.

Fox, D. (2002). Antecedent-contained deletion and the copy theory of movement. *Linguistic Inquiry*, 33(1):63–96.

Fox, D. and Hackl, M. (2006). The universal density of measurement. *Linguistics and Philosophy*, 29(5):537–586.

Frank, M. C., Braginsky, M., Yurovsky, D., and Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, 44(3):677.

Frege, G. (1879). Begriffsschrift. In van Heijenoort, J., editor, *Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*. Harvard University Press, Cambridge, MA.

Frege, G. (1892). Über sinn und bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100:25–50.

Frege, G. (1893). *Grundgesetze der Arithmetik: Begriffsschriftlich Abgeleitet*, volume 1. H. Pohle.

Gagnon, M. and Wellwood, A. (2011). Distributivity and modality: Where 'each' may go, 'every' can't follow. In *Semantics and Linguistic Theory*, volume 21, pages 39–55.

Gaillard, W. D., Berl, M. M., Moore, E., Ritzl, E., Rosenberger, L., Weinstein, S., Conry, J., Pearl, P., Ritter, F., Sato, S., et al. (2007). Atypical language in lesional and nonlesional complex partial epilepsy. *Neurology*, 69(18):1761–1771.

Gajewski, J. (2002). L-analyticity and natural language. *Unpublished*.

Garvey, C. and Hogan, R. (1973). Social speech and social interaction: Egocentrism revisited. *Child Development*, pages 562–568.

Gathercole, V. C. (1986). The acquisition of the present perfect: Explaining differences in the speech of scottish and american children. *Journal of Child Language*, 13(3):537–560.

Gelman, S. A. (2009). Generics as a window onto young childrens concepts. In Pelletier, F. J., editor, *Kinds, Things, and Stuff: Mass Terms and Generics*, pages 100–120. Oxford University Press.

Gelman, S. A., Coley, J. D., Rosengren, K. S., Hartman, E., Pappas, A., and Keil, F. C. (1998). Beyond labeling: The role of maternal input in the acquisition of richly structured categories. *Monographs of the Society for Research in Child Development*, pages i–157.

Gelman, S. A., Taylor, M. G., Nguyen, S. P., Leaper, C., and Bigler, R. S. (2004). Mother-child conversations about gender: Understanding the acquisition of essentialist beliefs. *Monographs of the Society for Research in Child Development*, pages i–142.

Geurts, B., Katsos, N., Cummins, C., Moons, J., and Noordman, L. (2010). Scalar quantifiers: Logic, acquisition, and processing. *Language and Cognitive Processes*, 25(1):130–148.

Geurts, B. and Nouwen, R. (2007). 'at least' et al.: The semantics of scalar modifiers. *Language*, 83(3):533–559.

Gil, D. (1992). Scopal quantifiers: Some universals of lexical effability. In Kefer, M. and van der Auwera, J., editors, *Meaning and Grammar: Cross-Linguistic Perspectives*, pages 303–345. Mouton de Gruyter.

Gillette, J., Gleitman, H., Gleitman, L., and Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, 73(2):135–176.

Glass, L. (2021). The lexical and formal semantics of distributivity. *Glossa: A Journal of General Linguistics*, 6(1).

Gleason, J. B. (1980). The acquisition of social speech routines and politeness formulas. In *Language*, pages 21–27. Elsevier.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1(1):3–55.

Gomes, V., Huh, Y., and Trueswell, J. C. (2020). Not what you expect: The relationship between violation of expectation and negation. In *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*.

Haberman, J. and Whitney, D. (2011). Efficient summary statistical representation when change localization fails. *Psychonomic Bulletin & Review*, 18(5):855–859.

Haberman, J. and Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. In Wolfe, J. and Robertson, L., editors, *From Perception to Consciousness: Searching with Anne Treisman*, pages 339–349. Oxford University Press.

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

Hacquard, V. and Lidz, J. (2018). Children's attitude problems: Bootstrapping verb meaning from syntax and pragmatics. *Mind & Language*, 34(1):73–96.

Haggerty, L. C. (1930). What a two-and-one-half-year-old child said in one day. *The Pedagogical Seminary and Journal of Genetic Psychology*, 37(1):75–101.

Halberda, J. and Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, 44(5):1457.

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., and Germine, L. (2012). Number sense across the lifespan as revealed by a massive internet-based sample. *Proceedings of the National Academy of Sciences*, 109(28):11116–11120.

Halberda, J., Sires, S. F., and Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science*, 17(7):572–576.

Hall, W. S. and Tirre, W. C. (1979). The communicative environment of young children: Social class, ethnic, and situational differences. *Center for the Study of Reading Technical Report*, 125.

Halle, M. (2003). *From Memory to Speech and Back: Papers on Phonetics and Phonology, 1954-2002*. Walter de Gruyter.

Hart, B. and Risley, T. R. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Paul H Brookes Publishing.

Hart, B. and Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator*, 27(1):4–9.

Heibeck, T. H. and Markman, E. M. (1987). Word learning in children: An examination of fast mapping. *Child Development*, pages 1021–1034.

Heilmann, J., Weismer, S. E., Evans, J., and Hollar, C. (2005). Utility of the macarthurbates communicative development inventory in identifying language abilities of late-talking and typically developing toddlers. *American Journal of Speech-Language Pathology*.

Heim, I. and Kratzer, A. (1998). *Semantics in Generative Grammar*. Blackwell Oxford.

Hempel, C. G. (1937). Le problème de la vérité. *Theoria*, 3(2-3):206–244.

Hempel, C. G. (1945). Studies in the logic of confirmation (i.). *Mind*, 54(213):1–26.

Herburger, E. (2000). *What Counts: Focus and Quantification*. MIT Press.

Higginbotham, J. and May, R. (1981). Questions, quantifiers and crossing. *The Linguistic Review*.

Higginson, R. P. (1985). *Fixing-assimilation in Language Acquisition*. Washington State University dissertation.

Hodges, W. (2012). Formalizing the relationship between meaning and syntax. In Werning, M., Hinzen, W., and Machery, E., editors, *The Oxford Handbook of Compositionality*, Oxford Handbooks, pages 245–261. Oxford University Press.

Hoeksema, J. (1983). Plurality and conjunction. In *Studies in Modeltheoretic Semantics*, pages 63–84. De Gruyter.

Hornstein, N. (2002). A grammatical argument for a neo-davidsonian semantics. In Preyer, G. and Peter, G., editors, *Logical Form and Language*. Oxford University Press.

Hunter, T. and Lidz, J. (2013). Conservativity and learnability of determiners. *Journal of Semantics*, 30(3):315–334.

Hunter, T., Lidz, J., Odic, D., and Wellwood, A. (2017). On how verification tasks are related to verification procedures: A reply to kotek et al. *Natural Language Semantics*, 25(2):91–107.

Im, H. Y. and Halberda, J. (2013). The effects of sampling and internal noise on the representation of ensemble average size. *Attention, Perception, & Psychophysics*, 75(2):278–286.

Inhelder, B. and Piaget, J. (1958). *The Growth of Logical Thinking from Childhood to Adolescence: An Essay on the Construction of Formal Operational Structures*, volume 22. Psychology Press.

Ioup, G. (1975). Some universals for quantifier scope. In *Syntax and Semantics*, volume 4, pages 37–58. Brill.

Izard, V. and Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, 106(3):1221–1247.

Jacobson, P. (1999). Towards a variable-free semantics. *Linguistics and philosophy*, pages 117–184.

Jacobson, P. I. (2014). *Compositional semantics: An introduction to the syntax/semantics interface*. Oxford University Press.

Jenkin, Z. (2020). The epistemic role of core cognition. *Philosophical Review*, 129(2):251–298.

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

Kahneman, D., Treisman, A., and Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, 24(2):175–219.

Kaufman, J., Csibra, G., and Johnson, M. H. (2005). Oscillatory activity in the infant brain reflects object maintenance. *Proceedings of the National Academy of Sciences*, 102(42):15271–15274.

Keenan, E. L. and Stavi, J. (1986). A semantic characterization of natural language determiners. *Linguistics and Philosophy*, 9(3):253–326.

Knowlton, T., Hunter, T., Odic, D., Wellwood, A., Halberda, J., Pietroski, P., and Lidz, J. (2021a). Linguistic meanings as cognitive instructions. *Annals of the New York Academy of Sciences*, n/a(n/a):xx–xx.

Knowlton, T. and 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*, volume 45, pages 399–412.

Knowlton, T., Pietroski, P., Halberda, J., and Lidz, J. (accepted). The mental representation of universal quantifiers. *Linguistics and Philosophy*.

Knowlton, T., Pietroski, P., Williams, A., Halberda, J., and Lidz, J. (2021b). Determiners are "conservative" because their meanings are not relations: Evidence from verification. In *Semantics and Linguistic Theory*, volume 30, pages 206–226.

Knowlton, T., Wong, A., Halberda, J., Pietroski, P., and Lidz, J. (2018). Different determiners, different algorithms: Two majority quantifiers in cantonese bias distinct verification strategies. *Poster presented at The 31st CUNY Conference on Human Sentence Processing, UC Davis*.

Kratzer, A. (1996). Severing the external argument from its verb. In *Phrase Structure and the Lexicon*, pages 109–137. Springer.

Krifka, M. (1992). Thematic relations as links between nominal reference and temporal constitution. In Szabolcsi, A. and Sag, I., editors, *Lexical Matters*, pages 29–53. Cambridge University Press.

Krueger, L. E. (1984). Perceived numerosity: A comparison of magnitude production, magnitude estimation, and discrimination judgments. *Perception & Psychophysics*, 35(6):536–542.

Kuczaj II, S. A. (1977). The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior*, 16(5):589–600.

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

Lakoff, G. P. (1966). *On the Nature of Syntactic Irregularity*. Indiana University dissertation.

Landau, B., Smith, L., and Jones, S. (1998). Object shape, object function, and object name. *Journal of Memory and Language*, 38(1):1–27.

Landau, B., Smith, L. B., and Jones, S. (1992). Syntactic context and the shape bias in children's and adults' lexical learning. *Journal of Memory and Language*, 31(6):807–825.

Landau, B., Smith, L. B., and Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, 3(3):299–321.

Landman, F. (1989a). Groups, i. *Linguistics and Philosophy*, 12(5):559–605.

Landman, F. (1989b). Groups, ii. *Linguistics and Philosophy*, 12(6):723–744.

Landman, F. (2000). *Events and Plurality*. Kluwer.

Landman, F. (2012). *Events and Plurality: The Jerusalem Lectures*, volume 76. Springer Science & Business Media.

Larson, R. and Segal, G. (1995). *Knowledge of Meaning: An Introduction to Semantic Theory*. MIT Press.

Larson, R. K. (2021). Rethinking cartography. *Language*, 97(2):245–268.

Lasersohn, P. (1995). Plurality, conjunction, and events, vol. 55 of. *Studies in Linguistics and Philosophy*.

Lasersohn, P. (2021). Common nouns as modally non-rigid restricted variables. *Linguistics and Philosophy*, 44(2):363–424.

LaTerza, C. (2014). *Distributivity and Plural Anaphora*. University of Maryland dissertation.

Lepore, E. and Ludwig, K. (2007). *Donald Davidson's Truth-theoretic Semantics*. Oxford University Press.

Leslie, S.-J. (2012). Generics articulate default generalizations. *Recherches Linguistiques de Vincennes*, 41:25–44.

Leslie, S.-J. and Gelman, S. A. (2012). Quantified statements are recalled as generics: Evidence from preschool children and adults. *Cognitive Psychology*, 64(3):186–214.

Leslie, S.-J. and Lerner, A. (2016). Generic Generalizations. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2016 edition.

Levin, B. and Rappaport Hovav, M. (2005). *Argument Realization*. Cambridge University Press.

Lewis, D. (1975). Languages and language. In Gunderson, K., editor, *Minnesota Studies in the Philosophy of Science*, volume 7. University of Minnesota Press.

Liberman, A. M. and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1):1–36.

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

Lidz, J. (2020). From structuralism to cognitive science: Lila Gleitman's contributions to the study of language and learning. In Gleitman, L. and Lidz, J., editors, *Sentence First, Arguments Afterwards: Essays on Language and Learning*, pages 1–21. Oxford University Press.

Lidz, J., Gleitman, H., and Gleitman, L. (2003). Understanding how input matters: Verb learning and the footprint of universal grammar. *Cognition*, 87(3):151–178.

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

Lidz, J. and Perkins, L. (2018). Language acquisition. *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, 4:1–49.

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

Link, G. (1991). Plural. In von Stechow, A. and Wunderlich, D., editors, *Semantics: An International Handbook of Contemporary Research*, pages 418–40. De Gruyter.

Ludlow, P. (2002). LF and natural logic. In Preyer, G. and Peter, G., editors, *Logical Form and Language*, pages 132–168. Oxford University Press.

MacWhinney, B. (1991). The CHILDES language project: Tools for analyzing talk. *Journal of Speech, Language and Hearing Research*, 40:62–74.

MacWhinney, B. and Snow, C. (1990). The child language data exchange system: An update. *Journal of Child Language*, 17(2):457–472.

Marr, D. (1982). *Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information*. MIT Press.

Masur, E. F. and Gleason, J. B. (1980). Parent–child interaction and the acquisition of lexical information during play. *Developmental Psychology*, 16(5):404.

May, R. (1985). *Logical Form: Its Structure and Derivation*. MIT Press.

Mbwana, J., Berl, M., Ritzl, E., Rosenberger, L., Mayo, J., Weinstein, S., Conry, J., Pearl, P., Shamim, S., Moore, E., et al. (2009). Limitations to plasticity of language network reorganization in localization related epilepsy. *Brain*, 132(2):347–356.

McCune, L. (1995). A normative study of representational play in the transition to language. *Developmental Psychology*, 31(2):198.

Medina, T. N., Snedeker, J., Trueswell, J. C., and Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences*, 108(22):9014–9019.

Minai, U., Jincho, N., Yamane, N., and Mazuka, R. (2012). What hinders child semantic computation: Children's universal quantification and the development of cognitive control. *Journal of Child Language*, 39(5):919.

Montague, R. (1973). The proper treatment of quantification in ordinary english. In *Approaches to Natural Language*, pages 221–242. Springer.

Morisset, C. E., Barnard, K. E., Greenberg, M. T., Booth, C. L., and Spieker, S. J. (1990). Environmental influences on early language development: The context of social risk. *Development and Psychopathology*, 2(2):127–149.

Mostowski, A. (1957). On a generalization of quantifiers. *Fundamenta Mathematicae*, 44(1):12–36.

Musolino, J., Crain, S., and Thornton, R. (2000). Navigating negative quantificational space. *Linguistics*, 38(1):1–32.

Musolino, J. and Lidz, J. (2003). The scope of isomorphism: Turning adults into children. *Language Acquisition*, 11(4):277–291.

Musolino, J. and Lidz, J. (2006). Why children aren't universally successful with quantification. *Linguistics*, 44(4):817–852.

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, 17(2):357–374.

Nelson, K. E. (1989). *Narratives From the Crib.* Harvard University Press.

Newman, R. S., Rowe, M. L., and Ratner, N. B. (2016). Input and uptake at 7 months predicts toddler vocabulary: The role of child-directed speech and infant processing skills in language development. *Journal of Child Language*, 43(5):1158–1173.

Nicholas, J. G. and Geers, A. E. (1997). Communication of oral deaf and normally hearing children at 36 months of age. *Journal of Speech, Language, and Hearing Research*, 40(6):1314–1327.

Ninio, A., Snow, C. E., Pan, B. A., and Rollins, P. R. (1994). Classifying communicative acts in children's interactions. *Journal of Communication Disorders*, 27(2):157–187.

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

Odic, D., Pietroski, P., Hunter, T., Halberda, J., and 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).

Papafragou, A., Cassidy, K., and Gleitman, L. (2007). When we think about thinking: The acquisition of belief verbs. *Cognition*, 105(1):125–165.

Parsons, T. (1989). The progressive in english: Events, states and processes. *Linguistics and Philosophy*, 12(2):213–241.

Parsons, T. (2014). *Articulating Medieval Logic*. Oxford University Press.

Partee, B. H. (2018). Changing notions of linguistic competence in the history of formal semantics. In Rabern, B. and Ball, D., editors, *The Science of Meaning: Essays on the Metatheory of Natural Language Semantics*, pages 172–196. Oxford University Press.

Perkins, L. (2019). *How Grammars Grow: Argument Structure and the Acquisition of Non-basic Syntax*. University of Maryland dissertation.

Philip, W. (1994). *Event Quantification in the Acquisition of Universal Quantification*. University of Massachusetts-Amherst dissertation.

Piantadosi, S. T., Goodman, N., Ellis, B. A., and Tenenbaum, J. (2008). A bayesian model of the acquisition of compositional semantics. In *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*, pages 1620–1625. Cognitive Science Society.

Pietroski, P., Lidz, J., Hunter, T., and 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., and Halberda, J. (2011). Seeing what you mean, mostly. *Syntax and Semantics*, 37:181–218.

Pietroski, P. M. (2003a). Quantification and second-order monadicity. *Philosophical Perspectives*, 17:259–298.

Pietroski, P. M. (2003b). Small verbs, complex events: Analyticity without synonymy. In Antony, L. M. and Hornstein, N., editors, *Chomsky and his Critics*, pages 179–214. Wiley Online Library.

Pietroski, P. M. (2005). *Events and Semantic Architecture*. Oxford University Press.

Pietroski, P. M. (2017). Semantic internalism. In McGilvray, J., editor, *The Cambridge Companion to Chomsky*, pages 196–216. Cambridge University Press.

Pietroski, P. M. (2018). *Conjoining Meanings: Semantics Without Truth Values*. Oxford University Press.

Poeppel, D., Idsardi, W. J., and Van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493):1071–1086.

Preminger, O. (2020). Natural language without semiosis. *Unpublished; Slides available at https://omer.lingsite.org/talks-handouts/.*

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, 80(1-2):127–158.

Pylyshyn, Z. W. and Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, 3(3):179–197.

Rappaport Hovav, M. and Levin, B. (1998). Building verb meanings. In Butt, M. and Geuder, W., editors, *The Projection of Arguments: Lexical and Compositional Factors*, pages 97–134. University of Chicago Press.

Rasin, E. and Aravind, A. (2021). The nature of the semantic stimulus: The acquisition of *every* as a case study. *Natural Language Semantics*, 29(2):339–375.

Rescher, N. (1962). Plurality quantification. *Journal of Symbolic Logic*, 27(3):373–374.

Roberts, C. (1987). *Modal Subordination, Anaphora, and Distributivity*. University of Massachusetts-Amherst dissertation.

Roeper, T., Pearson, B. Z., and Grace, M. (2011). Quantifier spreading is not distributive. In *Proceedings of the Boston University Conference on Language Development*, volume 35, pages 526–539.

Rollins, P. and Trautman, C. (2011). Caregiver input before joint attention: The role of multimodal input. In *International Congress for the Study of Child Language (IASCL)*.

Rollins, P. R. (2003). Caregivers' contingent comments to 9-month-old infants: Relationships with later language. *Applied Psycholinguistics*, 24(2):221.

Rollins, P. R. and Greenwald, L. C. (2013). Affect attunement during mother-infant interaction: How specific intensities predict the stability of infants' coordinated joint attention skills. *Imagination, Cognition and Personality*, 32(4):339–366.

Romero, M. (2015). The conservativity of many. In *Proceedings of the 20th Amsterdam Colloquium*, pages 20–29.

Romoli, J. (2015). A structural account of conservativity. *Semantics-Syntax Interface*, 2(1):28–57.

Rondal, J. A. (1976). *Maternal Speech to Normal and Down's Syndrome Children Matched for Mean Length of Utterance*. University of Minnesota dissertation.

Sachs, J. and Nelson, K. (1983). Talking about the there and then: The emergence of displaced reference in parent-child discourse. *Children's Language*, 4:1–28.

Safir, K. and Stowell, T. (1987). Binominal each. In *Proceedings of the North East Linguistics Society*, volume 18, pages 426–450. ScholarWorks at UMass Amherst.

Sawyer, R. K. (2013). *Pretend Play as Improvisation: Conversation in the Preschool Classroom*. Psychology Press.

Scha, R. J. (1984). Distributive, collective and cumulative quantification. In *Truth, Interpretation and Information: Selected Papers from the Third Amsterdam Colloquium*, volume 2, pages 131–158. De Gruyter Mouton.

Schein, B. (1993). *Plurals and Events*, volume 23. MIT Press.

Schein, B. (2012). Event semantics. In Graff Fara, D. and Russell, G., editors, *The Routledge Companion to the Philosophy of Language*, pages 280–294. Routledge.

Scholl, B., Pylyshyn, Z., and Franconeri, S. (1999). When are featural and spatiotemporal properties encoded as a result of attentional allocation? *Investigative Ophthalmology & Visual Science*, 40(4):S797.

Scholl, B. J. and Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, 38(2):259–290.

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

Smith, L. and Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, 106(3):1558–1568.

Soderstrom, M., Blossom, M., Foygel, R., and Morgan, J. L. (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language*, 35(4):869–902.

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

Spelke, E. S. (2000). Core knowledge. *American Psychologist*, 55(11):1233.

Spelke, E. S. and Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, 10(1):89–96.

Spenader, J. and de Villiers, J. (2019). Are conservative quantifiers easier to learn? evidence from novel quantifier experiments. In *Proceedings of the 22nd Amsterdam Colloquium*, pages 504–512.

Sportiche, D. (1988). A theory of floating quantifiers and its corollaries for constituent structure. *Linguistic Inquiry*, 19(3):425–449.

Sportiche, D. (2005). Division of labor between merge and move: Strict locality of selection and apparent reconstruction paradoxes. *Unpublished*.

Steinert-Threlkeld, S., Munneke, G.-J., Szymanik, J., et al. (2015). Alternative representations in formal semantics: A case study of quantifiers. In *Proceedings of the 20th Amsterdam Colloquium*, pages 368–377.

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

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

Sugisaki, K. and Isobe, M. (2001). Quantification without qualification without plausible dissent. *University of Massachusetts Occasional Papers in Linguistics*, 25:97–100.

Suppes, P. (1974). The semantics of children's language. *American Psychologist*, 29(2):103.

Surányi, L. B. (2003). *Multiple Operator Movements in Hungarian*. Utrecht University dissertation.

Sweeny, T. D., Wurnitsch, N., Gopnik, A., and Whitney, D. (2015). Ensemble perception of size in 4–5-year-old children. *Developmental Science*, 18(4):556–568.

Syrett, K. and Musolino, J. (2013). Collectivity, distributivity, and the interpretation of plural numerical expressions in child and adult language. *Language Acquisition*, 20(4):259–291.

Syrett, K., Musolino, J., and Gelman, R. (2012). How can syntax support number word acquisition? *Language Learning and Development*, 8(2):146–176.

Szabolcsi, A. (1997). Strategies for scope taking. In Szabolcsi, A., editor, *Ways of Scope Taking*, pages 109–154. Springer.

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

Talmina, N., Kochari, A., Szymanik, J., et al. (2017). Quantifiers and verification strategies: Connecting the dots. In *Proceedings of the 21st Amsterdam Colloquium*, pages 465–473.

Tarski, A. (1956). The concept of truth in formalized languages. *Logic, Semantics, Metamathematics*, 2(7):152–278.

Tomaszewicz, B. (2011). Verification strategies for two majority quantifiers in polish. In *Proceedings of Sinn und Bedeutung*, volume 15, pages 597–612.

Trautman, C. H. and Rollins, P. R. (2006). Child-centered behaviors of caregivers with 12-month-old infants: Associations with passive joint engagement and later language. *Applied Psycholinguistics*, 27(3):447.

Tunstall, S. (1998). *The Interpretation of Quantifiers: Semantics and Processing.* University of Massachusetts-Amherst dissertation.

Valian, V. (1991). Syntactic subjects in the early speech of american and italian children. *Cognition*, 40(1-2):21–81.

Van Houten, L. J. (1986). The role of maternal input in the acquisition process: The communicative strategies of adolescent and older mothers with the language learning children. In *Proceedings of the Boston University Conference on Language Development*, volume 11, pages 143–150. ERIC.

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

Vogel, E. K., Woodman, G. F., and 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.

Ward, E. J., Bear, A., and 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.

Warren-Leubecker, A. (1982). *Sex Differences in Speech to Children.* Georgia Institute of Technology dissertation.

Warren-Leubecker, A. and Bohannon, J. N. (1984). Intonation patterns in child-directed speech: Mother-father differences. *Child Development*, pages 1379–1385.

Weist, R. M. and Zevenbergen, A. A. (2008). Autobiographical memory and past time reference. *Language Learning and Development*, 4(4):291–308.

Wellwood, A. (2020). Interpreting degree semantics. *Frontiers in Psychology*, 10:2972.

Wellwood, A., Gagliardi, A., and Lidz, J. (2016). Syntactic and lexical inference in the acquisition of novel superlatives. *Language Learning and Development*, 12(3):262–279.

Wellwood, A., Xiaoxue He, A., Lidz, J., and Williams, A. (2015). Participant structure in event perception: Towards the acquisition of implicitly 3-place predicates. *University of Pennsylvania Working Papers in Linguistics*, 21(1):32.

Westerståhl, D. (1985). Logical constants in quantifier languages. *Linguistics and Philosophy*, 8(4):387–413.

Westerståhl, D. (2019). Generalized quantifiers. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy.* Metaphysics Research Lab, Stanford University, winter 2019 edition.

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

Wiggins, D. (1980). 'Most' and 'all': Some comments on a familiar programme and on the logical form of quantified sentences. *Reference, Truth, and Reality*, pages 318–346.

Williams, A. (2015). *Arguments in Syntax and Semantics*. Cambridge University Press.

Williams, A. (2021). Events in semantics. In Stalmaszczyk, P., editor, *The Cambridge Handbook of the Philosophy of Language*, Cambridge Handbooks in Language and Linguistics. Cambridge University Press.

Wilson, B. and Peters, A. M. (1988). What are you cookinon a hot?: A three-year-old blind childs violationof universal constraints on constituent movement. *Language*, 64:249–273.

Winter, Y. (1997). Choice functions and the scopal semantics of indefinites. *Linguistics and Philosophy*, pages 399–467.

Winter, Y. (2000). Distributivity and dependency. *Natural Language Semantics*, 8(1):27–69.

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

Wolfe, J. M. (1998). Visual search. In Pashler, H., editor, *Attention*, pages 13–73. University College London Press.

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, 85(3):223–250.

Xu, F. and Carey, S. (1996). Infants metaphysics: The case of numerical identity. *Cognitive psychology*, 30(2):111–153.

Xu, Y. and Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends in Cognitive Sciences*, 13(4):167–174.

Yuan, S. and Fisher, C. (2009). Really? She blicked the baby? Two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science*, 20(5):619–626.

Zeijlstra, H. (2011). On the syntactically complex status of negative indefinites. *The Journal of Comparative Germanic Linguistics*, 14(2):111–138.

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