THE MANY MEANINGS OF *MANY*: A PRAGMATIC-SEMANTIC ACCOUNT OF AMBIGUOUS, VAGUE QUANTIFIERS WITH EXPERIMENTAL SUPPORT

by
Karen Kou Clothier

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
November 2019

Abstract

Vagueness and ambiguity in natural language pose a challenge for formal theories of linguistic meaning because by definition they challenge two core assumptions: (i) that formal meanings can be formulated within a classic two-valued logic; and (ii) that an efficient communicative system maintains one-to-one mappings from meaning representations to pronounced lexical items (Kennedy 2011; 2019). At least as far back as Partee 1988, *many, few, much* and *little* have been characterized as ambiguous and vague quantifiers. Furthermore, their status as generalized quantifiers has been challenged (Barwise & Cooper, 1981; c.f. Herburger 2000). This dissertation uses *many* as a test case to explore particular formal linguistic tools that capture the apparent ambiguity, and vagueness of this class of quantifiers, and in formal accounts of meaning more generally. This dissertation offers a pragmatic-semantic account of their vagueness and ambiguity, and maintains their status as generalized quantifiers. To do so, structured meanings (Krifka 1995; 2001) denote licensing questions in Discourse trees (Büring 2003; Ginzburg 1997), which in turn provide the abstract functional content of the comparison class to the anaphoric Focus operator (Rooth 1992; von Fintel 1994) which composes with a relative, proportional determiner and a gradable degree morpheme at LF (Romero 2015b). This account captures the Focus-dependent interpretations of utterances containing determiners like *many*, whilst maintaining a single conservative, relative, proportional lexical entry for *many,* from which other senses are pragmatically derived. Two critical claims are made in this dissertation: (i) ambiguity stems from the interpretation of a *many*-utterance dependent on Focus-marking; and (ii) vagueness stems from the gradable, relative and proportional compositional semantics of *many*. Evidence from acceptability judgement experiments where the context and prosodic emphasis were manipulated failed to find clear evidence that Focus-marking alone can categorically disambiguate *many*. However, evidence from two speeded truth-verification experiments supports the denotation posited here. Thus, this work provides insight into formal mechanisms that can capture vagueness without abandoning a classic logic, and adds to the growing literature that emphasizes the importance of pragmatic and non-linguistic reasoning for linguistic production and comprehension, while leaving open a fully articulated account of the ambiguity of *many*.

Committee: Justin Bledin, Geraldine Legendre, Justin Halberda, Barbara Landau, Kyle Rawlins

Acknowledgements

The path that brought me to Johns Hopkins, to the Cognitive Science Department, working on this dissertation is crowded with people that have pushed me, guided me, taught me, supported me, and loved me, and I owe each of them a unique debt of gratitude. I attempt to describe these debts below, knowing full well there were numerous others who in turn guided and supported them, and that words alone cannot express my full gratitude to many of these individuals.

The work that underpins this dissertation never would have been started without the generosity and guidance of Kyle Rawlins and Barbara Landau. Kyle agreed to advise me, a novice semanticist, thereby enabling my continued place in the program. Without his patient tutelage and generous support, I never would have made it to this point. Kyle's ability to pinpoint an explanatory or analytical weakness and hint at solutions has helped me learn how to think more like a linguist, a skill I hope I never lose. I came to Barbara with a vague (pun intended) idea and she patiently listened and questioned as I refined my ideas. The great convener, Barbara excels at bringing ideas and people together. Her insightful, thoughtful leadership has taught me how to think both broadly and deeply, like an experimentalist. I am also grateful to Justin Halberda for generously providing his time and expertise in the late stages of the project, helping me create well-designed experiments to test my hypotheses. I stand in awe (and envy) at his ability to visualize data in his mind!

Around this dream team of advisors, I also thankfully built a crack team of cheerleaders, upon whom I selfishly placed the burden of maintaining my mental health, confidence and motivation throughout this monstrous mission. Thankfully they all proved more than up to the challenge. Thank you to Jane Lutken for being the best listener and empathizer; thank you to Alexey Krichtal for constantly surpassing my expectations of patience and empathy, and also feeding and loving me even when I was at my worst. Thank you to Celia Litovsky, Natalia Talmina, Ayushi Pandey, Sadhwi Srinivas, Najoung Kim and Emory Davis for being so distracting and fun! Finally, thank you to my parents for believing in me (ahem, not believing my threats to give up).

Last but not least, I owe every ounce of my personal ambition, conviction and stubborn independence to my grandparents Wendell and Yoko Hardy. My grandfather never let me think a thought or take an action without interrogating my motivations and influences. He, and our shared obsession with Star Trek, have instilled in me my devotion to empirical truth, equality, and rationality. He made me the scientist I strive to be. My grandmother was the strongest, most determined, talented, intelligent, ambitious, hard-working, ingenious woman I have ever met and she is my role model even in death. I believe the tenacity and creativity that I relied on to see this dissertation through to completion comes mostly from her, and I am truly thankful she was my Grammy.

Dedication

To Yoko Hardy

You made me stronger

Contents

List of Tables

List of Figures

1. Introduction

We can use natural language to equivocate, communicate uncertainty, play with double entendre, and speak in vague generalities. But such uses of language pose hard problems for theorists attempting to capture logically precise, binary truth-conditions. Moreover, theorists' understanding of how best to formally represent ambiguity and vagueness could have practical implications for the ever-increasing importance for Natural Language Processing and human-machine interaction. This dissertation posits a denotation for the supposedly ambiguous, vague quantifier *many* that is neither ambiguous nor vague. Instead ambiguity and vagueness are derived from the interaction of the semantic specification of meaning and the pragmatic contexts in which it is uttered.

This dissertation tackles the issues of ambiguity and vagueness using *many* as an illustrative case study. *Many* has been characterized as an ambiguous and vague quantifier, along with *few*, *much*, and *little* (Cohen 2001; Herburger 2000; Kennedy 2011; 2019; Partee 1988; Romero 2015a, 2015b; Solt 2011; 2015)[1]. Vagueness and ambiguity pose important theoretical challenges to standard semantic theories of logically specified truth-conditions because they raise questions about the validity of using classical binary logic in the specification of truth-conditions (c.f. Kennedy 2011; 2019; Parsons 1973;

[1] This dissertation focuses on the quantificational determiner uses of *many*. However, *many* appears in other semantically and syntactically distinct uses, e.g. as an adjective:

i. The many students who attended the lecture…
ii. There were too many students.

It remains for future work to determine whether the account developed here can be extended to these other uses, modulo semantic types (but see Solt 2015 for a promising analysis that bears a promising similarity to the approach adopted here, suggesting that this may be feasible). However, Solt 2011 suggests that due to their distinct distribution these uses warrant distinct semantic types and therefore may not be amendable to this analysis.

Solt 2011 a.o.), the nature of knowledge representation and how it relates to judgements about vague predicates, and why we are even inclined to make judgments on seemingly boundary-less meanings (Fara 2000; Kennedy 2011; 2019). However, given their prevalence in natural language, developing a fully articulated, formally specified account of the pragmatic and semantic knowledge and the cognition that supports the comprehension and production of ambiguous, vague quantifiers will provide valuable theoretical insight into principled ways of capturing ambiguity and vagueness with current, well-established linguistic theory, and how interfaces with other conceptual and perceptual systems affect linguistic judgements. In short, a better understanding of ambiguous, vague quantifiers could strengthen researchers' ability to investigate and understand a core question in cognitive science: how humans produce and comprehend language, especially in cases of ambiguity, vagueness, ignorance and/or uncertainty.

This dissertation investigates ambiguous and vague quantifiers, exemplified by *many*, in an attempt to address the following issues.

Theoretical Questions:

What is the best way to capture the vagueness and ambiguity of *many* and similar quantifiers?

a. Is the ambiguity of *many* the result of distinct denotations? If so, what are they? If not, why does *many* appear ambiguous?

b. Is the vagueness of *many* the result of something encoded directly in the denotation? If so, what is the nature of this representation? If not, what is the denotation of *many*, and how and why does it appear vague?

To preview the account developed in this dissertation, the apparent ambiguity of *many* arises because of differences in information structure in distinct contexts of use that change the contents of a contextual parameter, the comparison class, in the composition of *many*-utterances. The contents of this comparison class can be transmitted either linguistically (as in the contextualizing paragraphs used in the experiments in **Chapter 3**) or perceptually (as in the dot arrays used in the experiments in **Chapter 4**). Different interpretations of *many*-utterances arise based on the structure and contents of this comparison class. Therefore, there is only one underlying lexical denotation for *many*.

In addition, the account developed in this dissertation similarly claims that the apparent vagueness of *many* also arises because of differences in information structure in distinct contexts, which provide distinct contents to the comparison class of *many*. Therefore, *many* itself is not vague, rather the vagueness of *many*-utterances arises from the composition of a *many*-utterance with this contextual determined comparison class. Moreover, the verification of a *many*-utterance can be evaluated via representations of approximate magnitudes (the Approximate Number System; Dehaene 1997; c.f. Solt 2011). The lexical denotation of the quantifier relates the proportion of items under discussion to relevant alternatives in the comparison class, either linguistically provided (e.g. via paragraphs in **Experiments 1** and **2**) or perceptually provided (e.g. with displays of dots in **Experiments 3** and **4**). In short, the experiments described in the second part of this dissertation, investigate the following empirical questions:

Empirical Questions:

Experiments 1 and **2** investigate whether:

a. the apparent ambiguity of *many* does not necessitate an analysis that relies on multiple lexical denotations of *many*; rather can the various interpretations of *many* be shown to arise dependent on manipulations to the information structure – signaled via a particular prosodic contour – of the discourse in which *many* is uttered?

Experiments 3 and **4** investigate whether:

a. the denotation of *many* is relative and proportional, but critically the boundaries of its truth-conditions are clearly defined; that is, can the vagueness of *many* be attributed to a verification strategy which requires the comparison of multiple sets, the cardinalities or proportions of which are context dependent?

The sections that follow first describe how ambiguity is typically characterized and diagnosed by linguists. They then explore relevant examples of *many*-utterances and review previous accounts of *many*, to illustrate how and why *many* has justifiably been characterized as ambiguous, but ultimately why such an analysis is not fully justified by the data and potentially unnecessary. The same is done for vagueness. First, a description of how it has been characterized and diagnosed in the past is provided. The next section reviews relevant *many*-utterances and how previous accounts of *many* have capture this vagueness. Ultimately, this dissertation agrees with accounts that denote *many* with a comparison of proportions in a given context, which predicts precise truth-conditions, but predicts that the value of those conditions shift depending on context, thereby giving rise to an apparently indeterminant boundary, and hence judgements of vagueness.

1.1. Ambiguity

1.1.1. Characterizing and capturing ambiguity

This and the following sections present data and review the literature concerning ambiguity and vagueness along with extant accounts of ambiguity and vagueness in quantifiers like *many*. These data illustrate the various ways in which these so-called ambiguous, vague quantifiers can be interpreted and how they differ from other kinds of quantifiers.

Ambiguity, generally speaking, is the case in which a lexical item can be mapped to multiple, different meanings – following the standard assumption in semantics that denotations are mappings between lexical items and some meaning (Kennedy 2011; 2019; a.o.). Kennedy 2011 and 2019 reviews two widely used diagnostics for this kind of semantic ambiguity.

The first diagnostic, the contradiction test, is exemplified in (1) and (2), both adapted from Kennedy 2019.

(1) ?? Lily's mother is Sterling's father's sister. Since Lily is a girl, she is Sterling's cousin, but not his cousin.

(2) Sterling's cousin is funny, but she isn't funny.

In English, (1) is read as a contradiction. Since English does not mark gender on the word *cousin*, there is only one meaning of *cousin*, resulting in the oddness of (1). Therefore, *cousin* is not semantically ambiguous in English. In contrast, there is an interpretation of (2) that is not a contradiction – critically relying on the two meanings of *funny* (e.g. 'odd' and 'humorous'), demonstrating that *funny* is semantically ambiguous in English (i.e. it has more than one meaning).

The second diagnostic, the identity of sense test, is exemplified in (3) and (4).

(3) Sally swept the tile, but not the carpet.

(4) ?? Sally swept the competition, but not the chimney.

Parallelism in ellipsis essentially restricts the space of interpretation, necessitating a parallelism in sense. In these examples, (3) does not result in oddness because the same sense of *sweep*, i.e. *to use a broom*, can be maintained across both clauses. In contrast, (4) seems odd because the sense of *sweep* in the two clauses cannot be the same; in short, *to sweep a competition* means to win all the available prizes, whereas *to sweep a chimney* means to clear debris from a chimney. This kind of alternation, or regular polysemy, could in principle be captured by positing lexical ambiguity – a distinct denotation for each sense, capturing the specific arguments and their roles (e.g. patient vs. instrument). Alternatively, this alternation could be captured by positing that these argument roles are semantically underspecified – so that either an instrument or a patient can compose with one and the same lexical entry for the verb *sweep*. The latter approach maintains the intuitively close relationship between these various senses, i.e. whether the verb describes the physical motion of removing debris from a surface with a tool, or the more figurative sense where the verb describes the act of removing opportunities for competitors to win. In contrast, the former captures the distinct senses in a transparent way.

Formally speaking, there are a number of ways in which ambiguity can be captured each of which have their relative strengths and weaknesses depending on the particular kind of ambiguity to be accounted for: structurally distinct logical forms (LFs) can be used to capture scope ambiguity and homonymy; type-shifting operations that derive distinct truth-conditions, or denotations, can be used to capture cases of polysemy; and positing underspecified lexical denotations that only gain full specification during composition or interpretation have been used to capture the relationship between homonymy and regular polysemy, as well as other devices like metaphor and metonymy (Kennedy 2011; 2019). Critically, in all these approaches, ambiguity is captured by a many-to-one mapping from representation (syntactic or semantic) to pronunciation; this maintains the standard semantic assumption, that truth-conditions are representations of meaning within a two-valued logic. As the data below show, *many* appears to be homophonous across at least four different interpretations, suggesting that distinct logical forms (LFs) or semantic types could be at play here. However, the data below also point to a critical role for prosodic emphasis in this ambiguity, which is cashed out as context sensitivity in **Chapter 2**.

1.1.2. Ambiguity of *many*

Given these diagnostics, we can consider why previous literature has characterized *many* as **ambiguous**. Initially, *many* does not appear to pattern like other ambiguous items, based on the results of the contradiction and identity of sense tests, as (5)-(6) illustrate.

(5) ?? Many Scandinavians have won the Nobel Prize, but not many Scandinavians have won the Nobel Prize.

(6) Many Scandinavians have won the Nobel Prize, but not Andorrans.

Recall that if the lexical item of interest gives rise to a contradiction in a coordinated structure as in (5), this is taken as evidence that that lexical item is not ambiguous, because if there was another meaning or sense available the contradiction would not arise. Since (5) is intuitively contradictory, this suggests that *many* may not be ambiguous in the same was as *funny* is in (2) above. Recall that if the elided material in the second clause of (6) has the same sense as the first clause, and does not give rise to oddness, this is taken as evidence that the item of interest is not ambiguous. Since (6) does not give rise to an odd judgment, this suggests that *many* is not ambiguous in the way *sweep* or *funny* are in (2) and (4) above. Therefore, the results of both these diagnostics suggest that the apparent ambiguity of *many* is not the result of accidental polysemy (e.g. 'ha-ha' *funny*, and 'odd' *funny*), or of regular polysemy (e.g. *sweep with a broom* or *sweep the competition*).

Despite the results of these diagnostics, previous accounts of *many* have described at least two different possible interpretations of *many*-utterances, and characterize this as a kind of ambiguity (Barwise & Cooper 1981; Partee 1988; Westerståhl 1985). To illustrate the data that gave rise to this ambiguity claim, consider whether the utterance in (7) is true or false.

(7) Many Scandinavians have won the Nobel Prize.

If we define Scandinavia as Iceland, Norway, Sweden, Finland and Denmark, then roughly, 26 million people live in Scandinavia. As of October 2019, 63 Nobel Prizes have been awarded to Scandinavians. Given this information, it hardly seems true to assert that a significant proportion of the Scandinavia's population has won the Nobel

Prize, especially considering that these same countries have collectively won 579 Olympic Gold medals over the years, as **Table 1-1** summarizes.

Table 1-1. This table shows total number of Nobel Prizes (of all time, as of October 2019) and total number of Olympic gold medals (of all time, as of Winter 2018), that each of the five Scandinavian countries have won.

	Nobel Prizes*	Olympic Gold Medals†	Population
Norway	13	188	5 mil
Sweden	30	202	9.5 mil
Finland	5	144	5.4 mil
Iceland	1	0	0.3 mil
Denmark	14	45	5.6 mil
total	63	579	

* according to https://en.wikipedia.org/wiki/List_of_countries_by_Nobel_laureates_per_capita
† according to http://www.medalspercapita.com

However, some people judge (7) to be true (e.g. Westerståhl 1985), but critically, on the basis of a different interpretation of the utterance. Instead of taking (7) to be asserting something about the kinds of prizes or international honors Scandinavians win, Westerståhl takes (7) to assert something about the kinds of people who win Nobel Prizes, i.e. the number of Scandinavians, compared to people from other regions, who have won the Nobel Prize. Under this so-called **reverse interpretation**, the relevant data would be the number of Scandinavian Nobel Prize winners compared to the number of winners from other regions, for example as shown in **Table 1-2**. According to the Nobel Foundation, a total of 950 laureates have been awarded since the first in 1901.[2] So given that and the information in **Table 1-2**, it would be unfair to say that Scandinavia has the

[2] https://www.nobelprize.org/prizes/facts/nobel-prize-facts/

largest proportion of laureates, but they certainly have more than some regions, and therefore one may[3] be willing to accept (7), under this interpretation.

Table 1-2. The number of Nobel laurates in Scandinavia compared to other relevant regions, for the purposes of illustration.

	Nobel Prizes*
Scandinavia	63
South America†	11
Western Europe‡	332

* Source: https://en.wikipedia.org/wiki/List_of_countries_by_Nobel_laureates_per_capita
† Brazil, Argentina, Peru, Colombia, Chile, Venezuela, Ecuador, Bolivia, Uruguay, Paraguay
‡ Belgium, France, Ireland, Germany, Luxemburg, Monaco, Netherlands, U.K., Andorra, Spain, Portugal

The key semantic difference between Westerståhl's 1985 reverse interpretation and what will be referred to here as the **regular interpretation** is the order in which the arguments appear to compose with the quantifier (what is referred to as the **regular-reverse ambiguity**). In the regular interpretation, the arguments appear to compose in the same order as the surface syntax (e.g. the argument structure would be *many(Scandinavians, win Nobel Prizes)*), whereas in the reverse interpretation, they seem to compose in the opposite order such that the restriction of the quantifier contains the predicate *have won Nobel Prizes*, and the scope contains *Scandinavians* (i.e. *many(win Nobel Prizes, Scandinavians)*). So, is *many* ambiguous, and if so, how? One critical hint to this puzzle is the observation that the reverse interpretation only arises, or is only felicitous, when *Scandinavians* is contrastively emphasized (Herburger 2000; see also Cohen 2001; Romero 2015b), equated to Focus by some, but not all, authors (Cohen

[3] Considering the known comparators (e.g. consider the judgement of (7) if Western Europe had not been included in **Table 1-2**, or if North America, with 408 laureates had also been included). This is relevant to the discussion of vagueness to follow.

2001; Herburger 1997; 2000; Partee 1988; Romero 2015a; 2015b; Solt 2011; 2015; c.f. Hendriks 2004; Tomaszewicz 2016)

Consider the result of the contradiction diagnostic again, now with a particular pattern of prosodic emphasis we can reach an interpretation of (5) that is not a contradiction: (8) (here boldface text marks the prosodic contour that indicates a Contrastive Topic, or Büring's 2003 B-accent).

(8) Many Scandinavians have won the **Nobel Prize**, but not many **Scandinavians** have won the Nobel Prize.

Specifically, (7) is not contradictory if we take it to mean that of the people who have won the Nobel Prize, many of them are Scandinavian, but of all Scandinavians, not many have won the Nobel Prize. So perhaps *many* is not semantically ambiguous, but rather pragmatically ambiguous; the meaning of a *many*-utterance changes depending on the context in which it is uttered. Since the **regular** and **reverse interpretations** arise with different, distinct prosodic contours, perhaps the apparent ambiguity in the order of the arguments of these two meanings can be derived from some property of prosody, or Focus, rather than positing two lexical denotations, or stipulating some type-shifting rule in order to account for this compositional difference. This is precisely the aim of the account that is developed in **Chapter 2** and for which experimental support is sought in **Chapter 3**. The rest of this section provides further evidence that *many*-utterances do indeed have either a regular or reverse interpretation depending on the information structure of the preceding discourse, signaled by emphasis, and that another kind of ambiguity (what is called the **cardinal-proportional ambiguity** here) arises depending on what the interlocutors know or want to know – i.e. the contents of that information structure. This evidence further points to the disambiguating role of information structure – the underlying driver of prosodic emphasis, which is claimed to be sensitive to the information structure of the preceding discourse; emphasis is one marker of Focus, which signals new or newer information (Büring 2003, Schwarzschild 1999). Therefore, the following section briefly describes the theoretical notion of information structure and how it gives rise to both Focus-sensitivity and prosody.

To further grasp the way in which *many* is ambiguous, consider the data in (9) (adapted from Romero, 2015b):

(9) **Context**: Mary, John and Adam have a friendship predicated mainly on reading obscure Scottish authors. After each of them reads a book they place it on their bookshelf, the contents of which are as follows:

Mary: 6 books by Douglas; 3 books by Keath; 6 books by Hings; 3 books by McDawn; 1 book by McFire

John: 3 books by Douglas

Adam: 2 books by Douglas

After inspecting the contents of each of their bookshelves without knowing how many books each author has written in total, you proclaim:

a. **Mary** has read many/some books by Douglas.

b. Mary has read many/some books by **Douglas**.

According to Büring 2003, this Contrastive Topic (CT) prosody (again marked with boldface) gives rise to an interpretation that makes reference to other relevant alternatives in the context. Thus, under the intended interpretation of the utterances in (9a) what is at issue is the number of books by Douglas that Mary has read, compared to John and Adam. In contrast, what is at issue in the utterances in (9b) are the number of books by Douglas that Mary has read, compared to books by other Scottish authors. Taking this into consideration then, the intuitive meaning of the utterance in (9a) containing *many* would be paraphrased as "Out of Mary, John and Adam, the number of books by Douglas read by Mary is sufficiently more than that read by John or Adam (leaving an in depth discussion of what 'sufficiently more' would be for **Chapter 2**). Similarly, the meanings of the *many*-utterance in (9b) would be "Of all the books by different authors read by Mary, the number by Douglas is sufficiently more than the number of other books by other Scottish authors she has read." For comparison, consider the intuitive meaning of *some* in (9). It is fairly similar, except that 'sufficiently more' would be replaced with 'at least one' (c.f. Barwise & Cooper 1981). Critically however, because the meaning of *some* itself does not contain any relative comparison, the meaning (i.e. the truth-value) of the utterance does not change depending on emphasis. Thus, given the data in (9), the meaning of *many,* but not *some,* appears to differ depending on which alternatives are at issue, i.e. based on what is emphasized in the utterance. This an illustration of the regular-reverse ambiguity.

However, another kind of ambiguity is exemplified if you now learn the total number of books that each author has written, assuming your goal is to convey something meaningful about this information. If we again consider the same utterances in this slightly different context, we find a different pattern of judgements:

(10) Douglas has written 60 books, Keath 4, Hings 6, McDawn 3, McFire 2

a. **Mary** has read many/some all books by Douglas.

b. Mary has read (F) many/some books by **Douglas**.

Now, the intuitive meaning of the *many*-utterance in (10a) would be paraphrased as "Out of Mary, John and Adam, the proportion of books by Douglas that Mary has read is sufficiently larger than the proportion read by John or Adam." Similarly, the intuitive meaning of (10b) would be paraphrased as "Of all the proportions of books by different authors read by Mary, the proportion of books by Douglas is sufficiently larger than the proportion of other books by other Scottish authors she has read." Given this new context, (10b) no longer seems true since Mary has only read 1/6 of all the books that Douglas has written, but she has read all four books that Keath has written, half of McFire's oeuvre, and so forth. However, while this new contextual information changed intuitions about the truth-conditions for *many*, those for *some* do not appear to have changed (i.e. both (10a) and (10b) are true so long as Mary has read more than one book by Douglas, regardless of what that is compared to). This contrast illustrates one way in which ambiguous, vague quantifiers like *many* differ from first-order, logical quantifiers like *some*. This difference in the pattern of truth-values between the data in (9) and (10) illustrate the cardinal-proportional ambiguity.

To summarize, *many* and related quantifiers are at least four-ways ambiguous: the so-called **regular** ((9a) & (10a)) and **reverse interpretations** ((9b) & (10b)) arise depending on whether a phrase inside the *many*-phrase is emphasized (the regular interpretation), or whether a phrase outside the *many*-phrase is emphasized (the reverse interpretation)

(Cohen 2001; Romer 2015a; 2015b; but see Herburger 2000; 1997; Partee 1988; Westerståhl 1985 for alternative accounts). More generally, this regular-reverse ambiguity appears to depend on what is at issue, e.g. the number of books read by each member of the group, or all the books written by a particular author. The ambiguity between a **cardinal** (as in (9)) and a **proportional** (as in (10)) arises depending on what the interlocutor knows about the context of the utterance. So, in (9a) and (9b) when the speaker does not know the cardinality of the set being quantified (all the books by Douglas, here; what is referred to as the **quantified set** here after), no proportion can be calculated, and only reference to known whole cardinalities can be made. Once the complete, possible cardinality of the quantified set is known, then a proportion can be calculated, as illustrated in (10), resulting in the so-called **proportional interpretation**. Distinct lexical denotations have been posited to capture each of these meanings at various points in the literature.

However, the results of the diagnostics from the beginning of this section and the data in (9) and (10) suggest that the different meanings of *many* may not arise from lexically distinct denotations or via type shifting rules, as some accounts of this ambiguity would posit (e.g. Herburger 19997; 2000; Partee 1988; see Kennedy 2011; 2019; Parsons 1973 for the theoretical import of such an account). Instead **Chapter 2** argues (following others, e.g. Cohen 2001; Romero 2015a; 2015b), these facts suggest that the meanings of *many* can be principally and predictably derived based on the pragmatic factors of the particular discourse context in which they are uttered, implying that *many* is not actually lexically ambiguous, but rather pragmatically ambiguous.

1.2. Pragmatically disambiguating ambiguity: Information structure of discourses, Prosody, and Licensing Questions

As described above, the regular-reverse ambiguity of *many*-utterances arises depending on which element(s) of the utterance are emphasized. This is the distinction between the **regular interpretations** in (9a) and (10a) and the **reverse interpretations** in (9b) and (10b). Westerståhl 's (1985) classic example is repeated in (11) (see also Cohen 2001; Herburger 1997; 2000; Romero 2015a; 2015b)

(11) Many Scandinavians have won the Nobel Prize.

a. Many Scandinavians have won the **Nobel Prize**. *regular proportional*

<u>Paraphrase</u>: Of all the things that Scandinavians have won, the proportion that have won the Nobel Prize in Literature is larger than the proportion of them that have won other honors.

b. Many **Scandinavians** have won the Nobel Prize. *reverse proportional*

<u>Paraphrase</u>: Of all the people that have won the Nobel Prize, the proportion of them that have been Scandinavians is larger than the proportion of winners from other regions.

In English, the particular emphasis pattern that gives rise to this difference in meaning has been identified as Contrastive Topic, which despite its name, is a type of Focus (Büring 2003; Rooth 1992; von Fintel 1994; Wagner 2012). The general consensus in the literature is that Focus is used to mark new information in a discourse and introduces or raises the salience of alternatives to the Focused element (see Rooth 1992; Schwarzschild 1999; von Fintel 1994; but see also Krifka 1995). Focus is typically realized in English[4] with a particular kind of prosodic emphasis (Büring's 2003 A- and B-accents). Thus, the emphasis, or Focus-marking, that appears in utterances like (11a) can only be felicitously uttered in a particular context where the new information being contributed by the utterance concerns what kinds of internationally recognized honors (i.e. relevant, reasonable alternatives to the Nobel Prize) Scandinavians have won. In contrast, the emphasis, or Focus-marking, that appears in utterances like (11b) can only be felicitously uttered in a particular context where the new information being contributed by the utterance concerns who (i.e. relevant, reasonable alternatives to Scandinavians) has won the Nobel Prize. If this information is not new or salient, then the utterance will be infelicitous in the context, i.e., the alternatives introduced by Focus play an integral role in the felicity and interpretation of the utterance.

Other more general factors also affect felicity, and therefore may also be described in terms of information structure. Roberts' (1996/2012) influential theory of relevance, provides a useful framework for describing and explaining why certain utterances are judged felicitous or infelicitous in certain contexts. Roberts introduces the notion of Question Under Discussion (QUD) and stipulates that an utterance is only felicitous if it is relevant to – i.e. answers – the QUD. To this notion of relevance, we may also wish to

[4] In other languages, Focus is realized with particles (e.g. Japanese; Kuno 1973) or scrambling (Hungarian; Kiss 2008), which provide hints at the syntactic behavior of Focus at LF and how it composes with the other elements of these utterances, as discussed in **Chapter 2**. (NB: Kiss' work is part of a larger debate about whether the properties of Focus and Topic stem from their syntactic behavior or their relationship to the discourse information structure. No part of the debate hinges on this line of argumentation, and so the pragmatic and semantic behavior of Focus are taken at face value.)

add resolvedness – or the degree to which a question answers or settles a question. van Rooy 2003 implements this as a utility function that serves to rank relevant answers in terms of how well they meet the goals of the interlocutors and other aspects of the context (see also Ginzburg 1995a & b). Resolvedness allows us to account for differences in felicity depending on the goals of the interlocutors with respect to, e.g. whether an answer containing an exact cardinality is expected or whether a *many*-utterance will be enough.

Combining these notions of relevance and resolvedness, we can posit that *many*-utterances are felicitous when they are relevant and resolving answers to the QUD in a given discourse context. Given this relationship, QUDs that give rise to a particular interpretation are referred to as **licensing QUDs**. For example, consider the two questions and their relevant, resolving answers in (12) and (13).

(12) A: "How many Scandinavians have won different honors?"

B: "Many Scandinavians have won the **Nobel Prize**." *regular interpretation*

(13) C: "How many people of each nationality have won the Nobel Prize?"

D: "Many **Scandinavians** have won the Nobel Prize." *reverse interpretation*

These answers are taken to be relevant and resolving, if we assume that the goals of A do not require an exact number as a response. Thus, B provides a felicitous response to the QUD posed by A in (12), because B is providing the kind of information being requested by A, at the appropriate level of precision; A learns that of the honors won by

Scandinavians, some significantly large proportion have won Nobel Prizes as opposed to other well-known accolades. Conversely, the answer provided by D in (13) would not be felicitous because it is providing information that does not answer A's question; thus, A's question in (12) is the licensing QUD for the regular, but not the reverse interpretation of *many*. Following the same reasoning, D's response in (13) is a relevant, resolving answer to C's question, because it offers information about the degree to which different nationalities win Nobel Prizes and that Scandinavians win a significantly large proportion of them, at the appropriate level of precision. Therefore, C's question in (13) can be thought of as the licensing QUD for the reverse, but not the regular, interpretation of *many*.

Discourse trees (D-trees; Büring 2003; Krifka 2001; see also Cohen & Krifka 2014) offer a formal way of modeling this information structure (the question-answer pairs). D-trees are tree-like structures that visually represent the hierarchical relations between licensing QUDs, and the relevant, resolving answers they license, as illustrated in **Figure 1-1**. One major theoretical innovation in this dissertation is to denote these licensing QUDs using **structured meanings**, explicitly linking the information structural content of a discourse to the context sensitive composition of a *many*-utterance (Aloni, Beaver, Clark, and van Rooij, 2002; Krifka 1995; 2001). Structured meanings, discussed in more depth in **Chapter 2**, are tuples containing a background function (e.g. $\lambda d\,.(\{|x\colon won(x,\ Scandinavians)|) \geq d\}$ in the tree depicted in **Figure 1-1**) and a restrictor set (e.g. *DEGREES*). By applying the function to the set, one can derive the denotations of the relevant, resolving answers to the question which this structured meaning denotes (e.g. $[\![d - many\ Scandinavians\ have\ won\ the\ Nobel\ Prize]\!] =$ $\lambda d'.\ (|\{x: Scandinavians(x)\} \cap \{x: win(\boldsymbol{NP}, x\}|: |\{x: Scandinavians(x)\}|) \geq d)$.

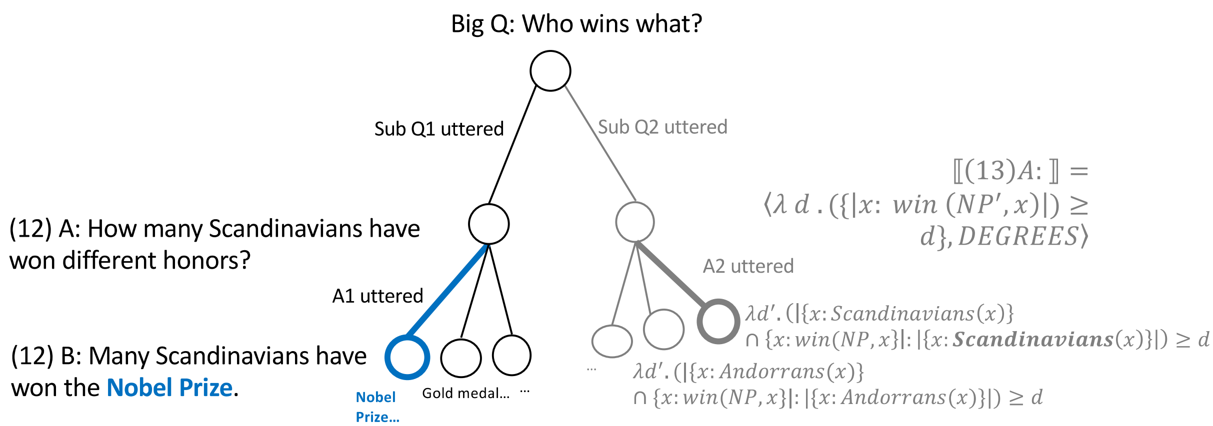


Figure 1-1. This D-tree illustrates how the structure of a discourse can be said to depend on, or be licensed by, the QUDs, either implicit or explicit, in the preceding discourse.

This formal structure of structured meanings also resembles that of Focus (i.e. a focus element, derived from a background and a restriction on that background, which can also be denoted with structured meanings, see Aloni, et al., 2002; Krifka 1995). This assumption makes the close relationship between licensing QUDs and Focus (marked or otherwise) explicit; in short, the unbound variable in the background function stipulates the element that can be marked with Focus. For example, in **Figure 1-1** the Big Question may be tacitly agreed by the interlocutors, but it is still thought of as a critical determinant of the tree's structure, because the sub-questions must each answer some part of the Big Question. On the left of the tree are the utterances from the dialogue in (12) in the text above, illustrating how each possible answer (each terminal node) must be a minimally contrasting answer to the directly dominating QUD in order to license the emphasis pattern particular to the regular interpretation. On the right branch of the tree, the same structure is illustrated, with the structured meaning denotation of the licensing QUD from the dialogue in (13) in the text above, and the denotation of the degree to which Scandinavians have won the Nobel Prize – the critical degree to be compared during the full composition of the *many*-utterance. This shows that by assuming that QUDs are denoted by structured meanings, the contents of the discourses' information structure can compose directly with a degree-based denotation of *many*, described in more detail in **Chapter 2**. Thus, the different interpretive possibilities of *many*-utterances, and possibly the prosody thereof, inextricably depend on the information structure of a particular discourse context.

To summarize, two distinct interpretations of a *many*-utterance can arise, exemplified by the distinct paraphrases in (11a) and (11b), depending on the information structure of the preceding discourse, each of which is signaled by a unique prosodic emphasis pattern, exemplified by the different pattern of boldface in (11a) and (11b). Formally, this occurs because of the strict felicity, or licensing, conditions on this emphasis, which is a marker of Focus. Because Focus must mark new information, Focus-marking allows us to deduce the goals of the interlocutors, what is already known and what is yet to be known in a discourse in which utterances like either (9a), (10a), (11a) or (9b), (10b), (11b) are made. **Chapter 2** will show that once the pragmatic information structure of the discourse is carefully characterized, one need only assume a single lexical entry for *many*; *many* is not ambiguous. The distinct interpretations of *many*-utterances can be predicted and derived based on the preceding discourse.

1.3. Vagueness

1.3.1. Characterizing and capturing vagueness

Vagueness, distinct from ambiguity, is the case in which the exact boundary between truth and falsehood is difficult to determine, not clearly defined, or shifts depending on factors like the context, the speaker etc. (Kennedy 2011; 2019). A necessary, although not sufficient, condition of vague predicates is their contextual variability – the truth-value of the predicate critically depends on the context. They also exhibit borderline cases – cases where it is not easily determined whether the predicate is true or false of an entity in some given context. For example, if a $5 cup of coffee is considered an expensive cup of coffee and a $1 cup of coffee is considered cheap (or not expensive), is a $4 cup of coffee expensive? Relatedly, vague predicates also give rise to the Sorites Paradox, illustrated in (14) (adapted from Kennedy 2019):

(14) **Premise 1:** A $5 cup of coffee is expensive.

Premise 2: Any cup of coffee that costs one cent less than $5 is an expensive cup of coffee.

Conclusion: Therefore, any free cup of coffee is an expensive cup of coffee.

The paradox lies in the intuitively false conclusion, drawn from two intuitively true premises (intuitively, $5 is expensive and so is $4.99, $4.98, and so on).

The fact that vague predicates give rise to the Sorites Paradox means that vague predicates raise some important questions about standard semantic assumptions about truth-conditional meaning and how knowledge and cognition interact with linguistic meaning. If there is some point in the iteration through successive values in Premise 2 at which we no longer accept Premise 2 (the so-called sharp boundaries claim), then how does this square with observations of borderline cases described above? More importantly, if there is no sharp boundary, why does a paradox still arise and how are we to capture the meaning of vague predicates like *expensive*, with a traditional binary logic (the so-called **semantic question**)? In addition to this question, Kennedy (2019; following Fara 2000) also argues that the Sorites Paradox raises questions about why we find it so hard to identify the point at which Premise 2 becomes false, even when we know all the relevant facts (the **epistemological question**); and it raises questions about why we are so willing to accept Premise 2 in the first place, i.e. why do vague predicates seem to have a wide range of truth-values for which both 'true' and 'false' judgements arise (tolerance), or seem boundary-less (the **psychological question**). These issues are revisited in **Chapter 2** after developing a pragmatic-semantic account of ambiguous, vague quantifiers.

There are two widely accepted formal mechanisms for capturing vagueness. Supervaluationist approaches (c.f. Fine 1975) assume sharp boundaries but introduce supertruth values (and superfalse) which are partial specifications of truth (or falsity). According to Fine's account, the edges of the sharp boundaries of extensions are arbitrary and infinitely variable, which is why we experience borderline cases and why we have such difficulty in rejecting the second premise of the Paradox. Gradable degree-based accounts (c.f. Hackl 2009; Rett 2014; Romero 2015a; 2015b; Solt 2015; a.m.o.) retain the classic binary logic, but introduce a new primitive into the ontology: degrees of type *d*.

Degrees are values on scales, the dimensions (units, upper, and lower bounds) of which are lexically or contextually specified. Accounts that posit ranges of truth or acceptability along these scales can account for tolerance and boundary-less judgements in the Sorites Paradox, these linguistic accounts on their own cannot address the epistemological or psychological questions that the Sorites Paradox raises. That is, they do not provide explanations of what we know about the meaning of a predicate if we cannot determine where/when it ceases to be true (even when we know, or think we know, all the relevant facts), and what we compute when we represent and process the meaning of vague predicates, i.e. what psychological factors make them tolerant and seemingly boundary-less? To address the epistemological question, an account must describe the nature of conceptual representations that linguistic meanings make reference to, or explain how the truth-conditions are calculated in such a way that gives rise to indeterminacy and imprecision. To address the psychological question, an account must describe the nature of the computational processes that operate over these conceptual or perceptual representations that give rise to judgements that seem tolerant or boundary-less. The account developed in **Chapter 2** attempts to answer both these questions. The next section briefly reviews evidence demonstrating the vagueness of *many*. Then, the subsequent section briefly previews the account developed in **Chapter 2**, sketching preliminary answers to both the psychological and epistemological questions.

1.3.2. Vagueness of *many*

Given this characterization of vagueness, let us return to the data in (8)/(9) and (10) (repeated here as (15) and (16)), and see why *many* has been characterized as **vague**, in addition to ambiguous (Cohen 2001; Kennedy 2011; 2019; Romero 2015a,b; Solt 2015, a.o.).

(15) **Context**: Mary has read 6, John 3, and Adam 2 of the 60 books that Douglas has written

a. Mary has read many books by Douglas.

(16) **Context**: Scandinavians have won 63 Nobel Prizes and 579 Olympic gold medals; Western Europe has won 332 Nobel Prizes and South America has won 11

a. Many Scandinavians have won Nobel Prizes.

Considering (15) and (16), we can see that there is no one, set cardinality or proportional threshold above which *many* is true and below which *many* is false; the truth-values depend in part on the alternatives in the context and nature of the entities being quantified. Recall that Mary has read six books by Douglas and six books by Hings. Under the cardinal interpretation of *many*, compared to her other two friends, she has read more than what they have managed and so intuitively this utterance is true, but this does not mean that six items is the boundary between *many* and *not-many*. However, if we altered the context of the utterance slightly so that we are no longer comparing Mary to John and Adam, but instead to either English majors (17b) or Engineering majors (17c), and we assume that English majors probably read a lot more obscure Scottish authors than Engineering majors, our intuitions about the truth-values of (17) may change. We may end up considering alternative individuals who have read hundreds of books in the first case, but individuals who have only read tens of books in the latter.

(17) Mary has read many books by Douglas,

a. …compared to John or Adam.

b. …compared to the English majors at University X.

c. …compared to the Electrical Engineering majors at University X.

Moreover, when we recall that Douglas has written 60 books, Keath 4, Hings 6, McDawn 3, McFire 2, and we now consider the difference in proportions of authors' oeuvres, our judgements also differ about whether six books count as *many*. This data exemplifies what Kennedy 2007 described as an "epistemic uncertainty about where we actually draw the line and metalinguistic resistance to treating highly similar objects differently relative to the property expressed by the [unmodified form of *many*]" (p. 42).[5]

In addition, referring to (18)-(20) we see that *many*-utterances in a given context can exemplify borderline cases, as in (19).

(18) If Mary has read 27 of Rowling's 30 books, Mary has read many books by Rowling.

(19) If John has read 15 of Rowling's 30 books, John has read many books by Rowling.

(20) If Adam has read 3 of Rowling's 30 books, Adam has read many books by Rowling.

Finally, *many*-utterances also give rise to the Sorites Paradox, as illustrated in (21).

[5] In natural, everyday use, there are, of course, instances where the cardinality of the quantified set may be guessed, estimated, or be relative to the particular entities being quantified or simply subjective, and be represented either correctly or incorrectly. This kind of indeterminacy, according to Kennedy (2007; 2011/2019; see also Zaroukian, 2013)), should be characterized as **imprecision** – a distinct, but interrelated phenomenon – which will not be discussed at length or formally distinguished in this dissertation, although *many* certainly exhibits such imprecision.

(21) **Premise 1:** Of all the Scandinavians who have won prestigious international prizes, 16 Scandinavian Nobel Literature Laureates count as many Scandinavians having won the Nobel Prize.

Premise 2: … 15 Scandinavian Nobel Literature Laureates…

Conclusion: … 0 Scandinavian Nobel Literature Laureates…

Therefore, *many* displays the widely accepted characteristics of vague predicates.

This dissertation follows Cohen 2001 and Romero 2015b by positing that the calculation of truth-conditions of ambiguous, vague quantifiers like *many* compares the proportion of quantified items out of all members of the quantified class (e.g. |Douglas' books read by Mary| : |all Douglas' books|) to the mean proportion of all relevant alternatives in the context (e.g. the mean of |Douglas' books read by John| : |all Douglas' books|, |Douglas' books read by Adam| : |all Douglas' books|, etc.). Moreover, values of these proportions are amenable to the Focus-marking discussed above. Specifically, when *Mary* is emphasized in (9), the alternative proportions would be as stated, but if *Douglas* was emphasized, the truth conditions would instead compare the proportion of Douglas' books read by Mary out of all Douglas' books to the mean proportion of each other relevant authors' books read by Mary. In short, the information structure specified by the preceding discourse and signaled by the particular Focus-marking determines the contents and form of the proportions being compared during the calculation of the truth-values for *many*. This provides an answer to Fara's 2000 epistemological question, we 'know' where the boundary between true and false is for *many*, but because it constantly shifts depending on subtle differences in the context, it is difficult to characterize and judge generally.

Moreover, the representation of these cardinalities or proportions themselves may also account for some of the imprecision exhibited by *many*. An important question is whether this vagueness should be represented in the linguistic denotation, or by some other cognitive system which interfaces with the linguistic system?

Following Solt (2009, 2011; 2015), and Kennedy (2007; 2011; 2019; a.o.), this dissertation claims that vagueness is not encoded in the lexical system, but can arise in part because the lexical system can interface with other cognitive systems to interpret lexical meaning denoted in the mental lexicon. The implementation that is pursued in this dissertation, following Romero (2015a & 2015b) and Solt (2011; 2015), posits that *many* relates sets of sets to a point on an infinitely gradable scale of degrees via a silent degree-morpheme, *POS*. Degree sets are ordered with respect to some dimension thereby forming a scale (Kennedy 2007; 2011; 2019), i.e. rational numbers for ambiguous, vague quantifiers like *many* (Romero, 2015b; Solt, 2009; 2011; 2015). That point, the degree of the **quantified set,** is then compared to an upper bound, or threshold derived from the relevant alternatives in the context (the 'neutral segment' in Romero's 2015b terms; derived via a function in the denotation of *POS* at LF). The utterance is true just in case the degree of the quantified set exceeds the neutral segment. Such a degree-based semantics, assumes that degrees are ontological primitives (like entities, truth-values, etc.), of type *d*. Assuming a degree-based approach provides a clear interface point between the linguistic system and the non-linguistic representation of approximate magnitudes. This provides one more source of vagueness and provides an answer to the Fara's (2000) psychological question: this cognitive system that represents approximate magnitudes cannot accurately or reliable discriminate small differences in magnitude. Therefore, small differences, like the difference between someone reading five or six books may not always be judged in reliably distinct ways.

To summarize, this account assumes that the truth-conditions for *many* clearly demarcate the boundary between true and false. The meaning of *many*-utterances appears indeterminant, tolerant, and boundary-less because the denotation of *many* entails the comparison of two proportional values, which themselves change depending on the context. Moreover, *many*-utterances may also appear vague because they can be verified with the noisy representations of approximate magnitudes. The next section briefly describes the nature of this system and this interface (see **Chapter 5** for more detail).

1.4. The interface between linguistic and non-linguistic representations

As hinted to above, language is not a completely isolated cognitive system. In order to produce and comprehend the meanings that language attempts to convey, the linguistic system must interface with a multitude of other cognitive systems. This is the crux of the Interface Transparency Thesis (ITT; Lidz, Pietroski, Halberda, & Hunter, 2011), which posits that there are instances when the nature of the linguistic representations strongly constrains, or fully determines the way in which non-linguistic cognitive systems are engaged in order to verify a linguistic utterance given a particular physical context. To interpret quantifiers, interlocutors clearly must make use of such interfaces, specifically with some system for representing quantities and individuating sets. As described below, this is the case when people are asked to consider the meaning of *most* (Lidz, et al., 2011; Pietroski, Lidz, Hunter & Halberda, 2009).

The experiments described in **Chapter 4** leverage what we know about the innate, non-linguistic system of representing approximate magnitudes, the Approximate Number System (ANS; Carey, 2009; Dehaene, 1997; Feigenson, Dehaene & Spelke, 2004) and individuating and tracking sets in parallel with visual attention and working memory (parallel individuation; Carey, 2009), in order to test the claim made above: that the denotation of *many* requires competent speakers of English to attempt to represent the approximate magnitude of each of the relevant, resolving answers to the QUD that licenses a particular interpretation of a *many*-utterance. The representations in the ANS are not modality specific; the ANS is a general-purpose quantity tracking or accumulating system that can represent, e.g. the approximate number of items in a set of physical objects, the approximate count of successive tones, accumulated seconds, etc. – but it does so imprecisely (Carey, 2009; Dehaene, 1997). As Lidz, et al., 2011 and Pietroski, et al., 2009 demonstrate, two well-characterized properties of these ANS representations can be leveraged to experimentally interrogate the proposed denotations of quantifiers. First, psychologists know that discriminating two magnitudes in the ANS follows a ratio-dependent performance curve such that discriminability decreases as the ratio between two magnitudes approaches 1:1, i.e. the difficulty of discriminating two magnitudes that stand in the same difference ratio to each other will be about the same (e.g. 5 vs. 10 and 10 vs. 20; Feigenson, Dehaene & Spelke, 2004). In addition, one's ability to rapidly and automatically individuate sets in parallel maxes out at about three sets (Carey, 2009; Halberda, Sires, & Feigenson, 2006).

The ITT (Lidz, et al., 2011) states that competent speakers of a language are biased towards verifying the truth-values of a given linguistic utterance in a way that is consistent with a "canonical specification of truth-conditions" (p.227) (i.e. a semantic denotation). Given the known limits on ANS representations, Pietroski, et al. (2009) reasoned that participants' patterns of responses to the three different kinds of trials exemplified in **Figure 1-2** would help determine which of two functionally equivalent verification strategies, in (22) and (23), accurately describe the way in which participants verified "Most of the dots are yellow."

(22) Greater-than [#{x: Dots(x) & Yellow(x)}, #{x: Dots(x) & ¬Yellow(x)}]

(23) One-to-one-plus [{x: Dot(x) & Yellow(x)}, {x: Dot(x) & ¬Yellow(x)}]

The participants' acceptance responses showed that they experienced increased difficulty assenting as the ratio between yellow and blue dots approached 1:1, regardless of whether the dots were scattered (as in **Figure 1-2 A**), paired off (as in **Figure 1-2 B**), or aligned and paired off (as in **Figure 1-2 C**); that is, participants' responses followed the ratio dependent-effects of discriminability, suggesting they used a magnitude based verification strategy supported by the ANS ((22)) instead a perceptual matching strategy ((23)), which would not have involved any representation of magnitude, and would therefore have been much easier and more accurate on displays like **Figure 1-2 B** and **Figure 1-2 C**.

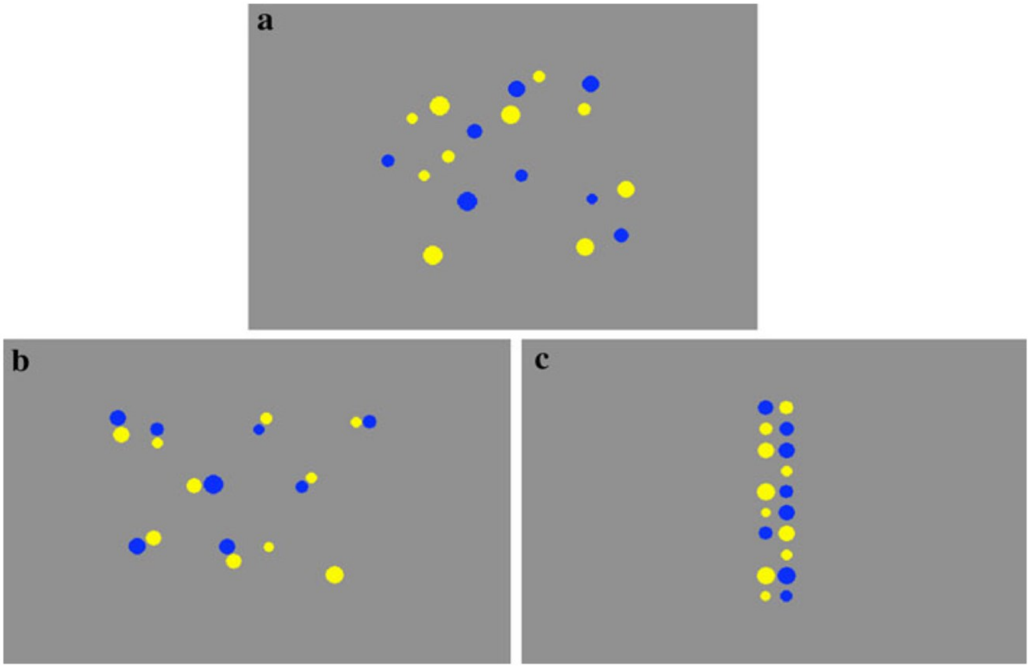


Figure 1-2. Example arrays from experiments designed to investigate the validity of the Interface Transparency Thesis with respect to the hypothesized verification strategy of *most* (adapted from Lidz, et al., (2011).

In a follow-up experiment, Lidz, et al. (2011), further probed the verification of *most*-utterances, by investigating whether *most* required only the individuation of the Quantified Set (e.g. yellow dots in these examples) as the strategy in (24) would suggest, or if each set on the trial needed to be individuated, as would be required in order to compute the strategy in (25). In addition to the ratio manipulation from Pietroski, et al. (2009), Lidz, et al. also manipulated the number different colored sets as illustrated in **Figure 1-3**, reasoning that given the known limits on parallel set individuation, performance would show decrements for trials with 3 or more sets if (25) was in fact the strategy participants used to verify "Most of the dots are yellow," but not (24).

(24) # of yellow dots > # of all dots - # of yellow dots

(25) # yellow dots > # of red dots + # blue dots

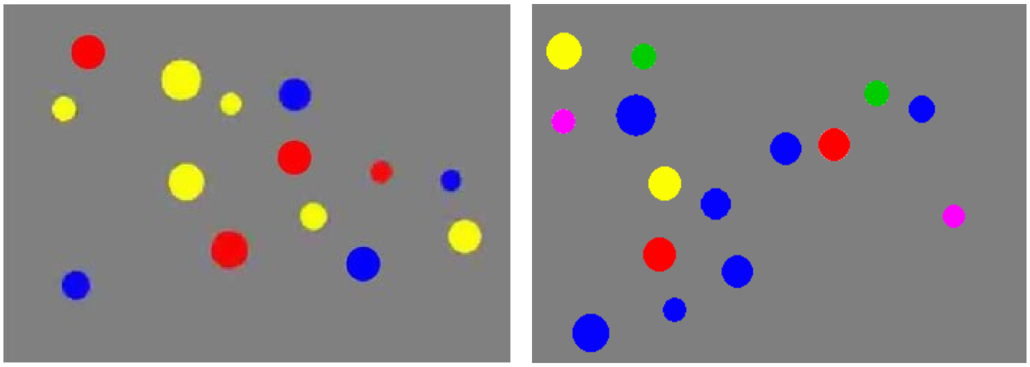


Figure 1-3. Example trials from Lidz, et al. (2011) showing a trial with just 3 different colored sets (left) and a trial with four different colored sets (right).

Participants in this experiment again showed the expected ratio-dependent decrements in discriminability, but no effects in their responses dependent on the number of different colored sets on a given trial. Lidz, et al. took this as evidence that the strategy adopted by participants did not require them to individuate each set on the trial, supporting the strategy given in (24) above.

The results of Pietroski, et al. (2009) and Lidz, et al. (2011) together show that in the context of rapidly, visually presented stimuli, what psychologists know about the ANS can help adjudicate the validity of proposed semantic denotations. The experiments described in **Chapter 4** similarly aim to leverage the ratio-dependent discriminability of ANS magnitude representations and limits on parallel set individuation in order to validate the relative, proportional denotation of *many* proposed in **Chapter 2**. In these experiments, both the ratio of the two magnitudes to be compared (e.g. the quantified set and the mean of all the sets) and the number of sets on a trial were manipulated in order to test whether participants who were asked to verify a *many*-utterance indeed attempted to individuate each set on the trial in order to compute the relevant mean, as the proposed denotation predicts.

1.5. Overview of chapters

This dissertation provides an account of the class of ambiguous, vague quantifiers exemplified by *many*, and provides some preliminary experimental evidence that: (i) the apparent ambiguity of *many* can in fact be reduced to pragmatic factors rather than distinct denotations, (ii) the apparent vagueness of *many* can also be attributed to pragmatic factors, rather than requiring some explicit change to the basic linguistic assumption of two-valued truth-conditions. The chapters that follow this introductory chapter are briefly summarized below.

Chapter 2 contains the main theoretical proposal in this dissertation. It provides a description of all the pragmatic and semantic components used to account for the context sensitive behavior of *many*. These components include Discourse-trees, licensing QUDs, structured meanings to denote licensing QUDs, information structure-dependent Focus, a Focus operator at LF, and a decompositional degree-based semantics for *many*. To summarize, this account claims that: the apparent cardinal-proportional ambiguity, and the regular-reverse ambiguity arise in a principled, predictable way depending on the information structure of the preceding discourse, or context; and the vagueness of *many* arises from the contents of the comparison class defined by, or determined, by the licensing QUD, i.e. the information structure of the preceding discourse.

Chapter 3 describes two experiments designed to test whether participants have access to the reverse interpretation via a particular prosodic contour (c.f. Büring 2003) i.e. whether distinct truth-conditions arise by virtue of emphasis (Focus) in the composition of *many*-utterances resulting either in the truth-values that arise from the emphasis pattern that gives rise to the regular interpretation or the reverse interpretation. To test this, paragraphs were designed with information about proportions of entities that made either the truth-conditions of the regular interpretation or the reverse interpretation of a *many*-utterance true. Participants were asked to select the best summary statement (a *many*-utterance with italics consistent with either the regular or reverse interpretations, e.g. "Many Scandinavians have won the *Nobel Prize*," or "Many *Scandinavians* have won the Nobel Prize."). In the first experiment, participants read these paragraphs and selected between the two *many*-utterances. In the second experiment participants heard audio recordings of the paragraphs and utterances. Assuming participants aim to summarize the paragraphs with utterances that evaluate to true, participants who only saw 'regular' paragraphs should overwhelmingly choose to summarize the paragraphs with the *many*-utterance with italics consistent with a prosodic contour that hypothetically gives rise to the regular interpretation. Conversely, participants who only saw 'reverse' paragraphs should overwhelmingly choose the *many*-utterance with italics consistent with a prosodic contour that hypothetically gives rise to the reverse interpretation. These results failed to show reliable differences in the rate at which participants chose either the regular or reverse *many*-utterance, regardless of the modality of presentation. However, participants were reliably more confident when they chose the reverse utterance after reading the 'reverse' paragraph. Moreover, a post-hoc Optimality Theory-based analysis, suggest that certain semantic constraints on the interpretation of the subjects of the predicate in the *many*-utterance may interact with the information structure, in determining which interpretations are available. These additional semantic factors are discussed, and further analyses taking them into account are sketched.

Chapter 4 describes two experiments designed to test whether participants who were asked to verify *many*-utterances take alternatives in the context into account, and compare the proportions of visually presented alternatives in a ratio- and set number-dependent way, as the relative, proportional denotation of *many* and the known limits of the ANS and parallel set individuation predict. To do so, these experiments were designed to force participants to perform this hypothesized comparison process using the noisy ANS representations of magnitude, by rapidly presenting (150-600ms) arrays with different colored dots (2-5 different colors depending on the trial) of varying sizes, either randomly scattered or grouped by color. The rapid presentation and large number of dots in each set (i.e. > 5) precluded counting, thereby forcing the use of the ANS and perceptual parallel set individuation, ensuring that both ratio- and set number-dependent effects in responses could be detected and used to infer the critical contextually determined relative proportionality of the proposed denotation for *many*. In a between-subjects design, half of all participants were asked to verify “Most of the dots were blue,” and the other half “Many of the dots were blue.” Previous work (Lidz, et al., 2011; Pietroski, et al., 2009) has shown that when participants are asked to verify *most*, the number of different colors does not alter performance. However, the account of *most* hypothesized and supported by these authors does not posit the same kind of relative proportional meaning as the current account of *many* does. In short, in the same way that the truth-values of “Many *Scandinavians* won the Nobel Prize,” critically depend on the numbers of other individuals of other nationalities who have won the Nobel Prize, the truth-values of "Many of the dots are blue," depends on the number of dots of other colors as well. Therefore, it was hypothesized that verification of *many*, compared to *most,* would show differences in performance (acceptance rate) depending on the number of different colored sets present on a given trial. Specifically, participants' ability and therefore willingness to accept the utterance should decrease as the number of different colored sets to be tracked increase, reflecting the perceptual systems' limited ability to accurately differentiate more than about three sets at a time (Halberda, Sires, & Feigenson, 2006), thereby causing a decrease in accuracy for participants asked to verify *many*, unlike *most*. Indeed, across both experiments, participants' willingness or ability to assent to "Many…" decreased as the number of dots in the Quantified Set and the average of the Alternative Sets approached equity, and as the number of different sets increased. Therefore, the results of these experiments support the claim that the underlying meaning of *many* requires the comparison of contextually defined quantities, and therefore the vagueness of *many*-utterances is attributable to this context-sensitivity.

Chapter 5 summarizes the claims of the theoretical proposal in **Chapter 2**, and the experimental results from **Chapters 3** and **4**. It then concludes by synthesizing all these results and drawing the following conclusions:

a. Positing a single proportional, degree-based denotation for *many*, accounts for the observed regular-reverse and cardinal-proportional ambiguities by assuming particular licensing discourses, the information structure of which can be captured by the structured meaning denotation of a particular licensing QUD. Experimental evidence and a further OT analysis, however, suggest that the role of information structure and the contents thereof are not the sole determinants of which of the possible interpretations of *many* arise. Rather, the interpretation of a *many*-utterance depends on constraints on semantic and pragmatic interpretation, syntactic configuration, and prosodic contouring. Therefore, assuming a single lexical entry for *many* may be a fruitful theoretical avenue to explore, but it will require much more work to define the full set of constraints that can adequately predict how and why a particular interpretation of *many* arises in a given utterance and discourse context.

b. Positing a relative, proportional, degree-based denotation for *many*, accounts for the observed vagueness because the values of the proportions to be compared depend on the content of each of the relevant sets in the context – since the contents and cardinality in each of those sets, and the number of relevant sets will change depending on the context, there is no one set boundary value for the truth-conditions of *many*. Experimental evidence supports this hypothesized context dependence by showing that participants do indeed attempt to individuate and estimate the cardinality of each set in a given perceptual context when asked to verify a *many*- but not a *most*-utterance.

2. The pragmatics and semantics of *many*

2.1. Overview of semantic accounts of *many*

As **Chapter 1** and previous semantic accounts of *many* (and related quantifiers like *few, much, little*) have observed, *many* is ambiguous, vague and most importantly context sensitive. Each of these characteristics has been explained in a variety of ways across different accounts. This section briefly reviews previous accounts' observations and analyses of these characteristics.

Ambiguity has appeared in several previous accounts of *many* in different forms. For example, Partee 1988 argued that there are in fact two distinct lexical entries of *many* – the **cardinal** and the **proportional** interpretation. The cardinal *many,* exemplified by the second conjunct in (1) and (2), according to Partee is denoted by (3).

(1) Many linguists are women iff there are many women linguists.

(2) If many men over 40 were there, then many men over 30 were there.

(3) $|A \ \cap B| > n$

These truth-conditions of cardinal *many* can be paraphrased as "Many (A, B) is true just in case the cardinality of the intersection of A and B is greater than some contextually defined numerosity, or cardinality." Thus, if the number of men over 40 counts as *many*, then so should the number of men over 30, since men over 30 are a proper subset of men over 40.

Partee 1988 notes that the denotation in (2) is intersective and persistent (illustrated by the acceptability of (1) and (2) respectively assuming the denotation in (3)). However, there are several formal properties of generalized quantifiers that the above analysis does not satisfy, thereby calling into question the status of *many* as a generalized quantifier within the logical framework of Barwise and Cooper (1981). Alternatively, Partee's 1988 proportional *many*, with the denotation in (4) does appear to have the desired properties (it is intersective and persistent), as exemplified by the unacceptability of (1) assuming the denotation in (4) and the acceptability of (5), when assuming the denotation in (4).

(4) $\frac{|A \cap B|}{|A|} \geq k\,; where\ k\ is\ a\ 'large' proportion$

(5) If many voters in Amherst will vote for the candidate, then many voters in the US will vote for the candidate.

Using these observations, Partee 1988 suggested that there are two syntactically and semantically distinct lexical entries for *many*: the proportional *many* that is a generalized quantifier ((4)), and the adjective-like cardinal *many* that is not a quantifier in the logical sense ((5)).[6]

Another potential ambiguity, however, in the determiner-like *many* was observed in Westerståhl 1985. Westerståhl 1985 considers cases like (6), and proposes a denotation as in (7), subsequently coined the **'reverse' proportional** denotation.

(6) **Context**: Of a total of 81 Nobel Prize winners in literature, 14 come from Scandinavia.

Many Scandinavians have won the Nobel Prize in literature.

(7) $\frac{|A \cap B|}{|B|} \geq k\,; where\ k\ is\ a\ large' proportion$

[6] While there may well be two semantically distinct types of *many*, issues surrounding the interpretation and use of the adjectival *many* are beyond the scope of this dissertation proposal. The remainder of this proposal will focus on the generalized quantifier-typed determiner *many*, which it is claimed has an inherently proportional meaning that can surface with a variety of interpretations.

If we interpret (6) with the denotation in (7), we can paraphrase the truth-conditions of (6) as "Many of the winners of the Nobel Prize in literature were Scandinavians." Hence, the name for this interpretation arises from the fact that the arguments appear to compose with the quantifier in the reverse order of how they appear in the surface syntax. Taken together these accounts propose three distinct denotations to capture three observed interpretations of *many*-utterances. *Many* appears to be quite ambiguous indeed.

Yet, the denotation proposed in Westerståhl 1985, like those in Partee 1988, questions the status of *many* as a generalized quantifier because it is not conservative (Barwise & Cooper, 1981; Romero 2015a). Moreover, these three distinct denotations do not offer a theoretically parsimonious account of the single lexical item *many*. Does this **reverse** interpretation have a denotation distinct from both (3) and (4) above – a three-way lexical ambiguity – or is (7) derived from (4) via a type-shifting rule, implying some kind of regular polysemy? The observation that *many*, *few*, *much*, and *little* are context sensitive has given rise to a number of accounts that all attempt to reconcile the apparent ambiguity and vagueness of these quantifiers. In these approaches, the context is assumed to contain information that specifies a comparison class of relevant alternatives to the quantified set (e.g. Americans and Andorrans compared to Scandinavians).

In addition, previous accounts have also noted the seemingly vague use of *many*. Here vagueness is taken to indicate the difficulty in pinning down exactly what counts as *many,* i.e. what is the cardinality of a set or proportion that counts as *many* (Kennedy 2011; 2019). Several accounts (e.g. Cohen 2001; Romero, 2015a; 2015b), have attributed this difficulty to the observed context sensitivity of *many*. In short, vagueness arises from the fact that interlocutors may disagree on exactly what groups of objects are being considered, i.e. what counts as relevant members of the comparison class (c.f. Rooth 1992; Szabolcsi 1986; 2010), which is a critical aspect of accounts such as Cohen 2001, Romero 2015b and, indeed, the account developed in this chapter. Others have also adopted a gradable, degree-based approach, positing that the vagueness of *many* can be captured with the gradient effects that arise from the continuous, infinite scale of rational numbers that form the degrees of *many* in such accounts (e.g. Romero 2015a; 2015b; Solt 2009; 2015)

To preview the account developed in this chapter, the context- (or Focus-) sensitivity of *many* is a critical characteristic. Specifically, context sensitivity, refers to the way in which the meaning of *many* – its truth-conditions and felicity conditions – observably change dependent on factors in the discourse context. For example, Cohen 2001, Herburger 1997 and 2000, and Romero 2015a and 2015b all acknowledged the role of phonological emphasis in disambiguating between the interpretations of *many*. In short, the underlying differences in discourses which license particular patterns of phonological emphasis give rise to different values used during the composition and interpretation of *many*. Under these accounts, there is no need to posit lexical ambiguity as proposed by Partee 1988 or Westerståhl 1985, instead this context sensitivity determines the different derived meanings.

In Cohen 2001, this phonological emphasis (marked by boldface in the examples below) gives rise to different alternative sets (c.f. (9a) and (11a) – the set that contains all the relevant alternatives to the focused phrase. For example, in (9) reasonable relevant comparisons to *cooks* might be *waiters*, *bartenders*, and so on, whereas the relevant comparisons for (11) might be *cooks that did not apply*. These comparison classes in turn give rise to meanings analogous to the regular proportional and reverse proportional meanings, (10) and (8), respectively. Thus, the meaning of the **reverse proportional interpretation** in (8) can be paraphrased as in (9), whereas the **regular proportional interpretation** in (10) can be paraphrased as in (11). The critical difference between these interpretations according to Cohen 2001 is the different comparison classes ((9a) and (11a)) that arise depending on different phonological emphasis – the same semantic denotation of *many* underlies both interpretations.

(8) Many **cooks** applied. *reverse interpretation*

(9) Many that applied were cooks.

a. Cooks compared to: {waiters, bakers, bartenders, etc.}

However, as Herburger 1997 and Cohen 2001, point out, this interpretation is actually only available with emphasis on *cooks*, thus (10) can only be paraphrased as (11) and not (9).

(10) Many cooks **applied**. *regular interpretation*

(11) Many cooks were cooks that applied.

a. Cooks that applied compared to {cooks that did not apply}

Cohen's proposal nicely reduces the ambiguity of *many* to pragmatic factors, motivating the account ultimately pursued in **Section 2.3**. In addition, his proposal of a relative, proportional denotation, given in (12) (where ϕ and ψ indicate the arguments to be saturated by each constituent phrase in the *many*-utterance), in part captures the vagueness of *many*: what counts as *many* shifts depending on the contents of the comparison class ($\cup A$) that composes with the other arguments in this denotation.

(12) $$\frac{|\psi \cap \phi|}{|\psi \cap \cup A|} > \frac{|\cup A \cap \phi|}{|\cup A|}$$

Therefore, context-sensitivity underlies both the ambiguity and vagueness of *many* in Cohen's approach. However, Cohen's account does not address the cardinal interpretation. Moreover, the denotation proposed in Cohen 2001 like Westerståhl 1985 is not conservative, again calling into question the status of *many* as a generalized quantifier.[7]

In another account, Romero 2015b derives a reverse interpretation of *many* from a conservative, generalized quantifier-typed denotation of *many*. To do so, Romero 2015b adopts a degree semantics approach, decomposing *many* into a generalized quantifier over degrees and a degree phrase, the denotations for which are given in (13) and (14), respectively:

(13) $$[\![many_{prop}]\!] = \lambda\, d_d . \lambda\, P_{\langle e,t \rangle} . \lambda\, Q_{\langle e,t \rangle} . (|P \cap Q| : |P|) \geq d$$

(14) $$[\![POS]\!] = \lambda\, Q_{\langle\langle d,t \rangle ,t\rangle} . \lambda P_{\langle d,t \rangle} . P \in Q . L_{\langle\langle dt,t\rangle\langle d,t\rangle\rangle}\, (Q) \subseteq P$$

This notion of *POS* is borrowed from accounts of gradable adjectives like *tall*, where the scale of degrees might be inches or feet. For *many*, the scale of degrees is assumed to be rational numbers. Thus, in adopting a degree-based approach, Romero 2015b captures the

[7] The debate concerning whether *many* should in fact be considered a generalized quantifier is beyond the scope of this dissertation. This dissertation, like Romero 2015 a and 2015b assumes that *many* is indeed a generalized quantifier, and should therefore have a conservative denotation.

vagueness of *many* by maintaining the relative proportional meaning of Cohen 2001, but also placing this comparison process on a continuous, infinite scale.

Romero 2015b also follows Cohen 2001 (and Herburger 2000) in using emphasis to derive the reverse interpretation of *many*, which she instantiates as association with Focus (Rooth 1992). This Focus can either associate with a phrase internal to the phrase in which *many* was base-generated, or with a phrase external to the base-generated *many*-phrase. In the former situation, the reverse interpretation arises as shown in the derivation in (15):

(15) Mary has read many books by **Douglas** $_{\textbf{F/CT}}$.

LF: [[POS C]] [1[Mary has read [t_1-many books by **Douglas** $_{\textbf{F/CT}}$]] ~C]

Comparison class derivation:

$$[\![C]\!] \subseteq \{\lambda d'. \left(\frac{|\{x: book(x) \wedge by(\boldsymbol{douglas}, x)\} \cap \{x: read(mary,\ x\}|:}{|\{x: book(x) \wedge by(\boldsymbol{douglas}, x)\}|} \right) \geq d,$$

$$\lambda d'. \left(\frac{|\{x: book(x) \wedge by(\boldsymbol{mcdawn}, x)\} \cap \{x: read(mary,\ x\}|:}{|\{x: book(x) \wedge by(\boldsymbol{mcdawn}, x)\}|} \right) \geq d \ldots\}$$

$$L([\![C]\!]) \subseteq \lambda d'. \left(\frac{|\{x: book(x) \wedge by(\boldsymbol{douglas}, x)\} \cap \{x: read(mary,\ x\}|:}{|\{x: book(x) \wedge by(\boldsymbol{douglas}, x)\}|} \right) \geq d$$

In contrast, when Focus associates with a phrase external to the base-generated *many*-phrase, the **regular interpretation** arises, as in (16):

(16) **Mary** $_{\textbf{F/CT}}$ has read many books by Douglas.

LF: [[POS C]] [1[**Mary** $_{\textbf{F/CT}}$ has read [t_1-many books by Douglas]] ~C]

Comparison class derivation:

$$[\![C]\!] \subseteq \{\lambda d'. \left(\begin{matrix}|\{x: book(x) \wedge by(douglas, x)\} \cap \{x: read(\boldsymbol{mary},\ x\}|: \\ |\{x: book(x) \wedge by(douglas, x)\}|\end{matrix}\right) \geq d,$$

$$\lambda d'. \left(\begin{matrix}|\{x: book(x) \wedge by(douglas, x)\} \cap \{x: read(\boldsymbol{john},\ x\}|: \\ |\{x: book(x) \wedge by(douglas, x)\}|\end{matrix}\right) \geq d \ldots\}$$

$$L([\![C]\!]) \subseteq \lambda d'. \left(\begin{matrix}|\{x: book(x) \wedge by(douglas, x)\} \cap \{x: read(\boldsymbol{mary},\ x\}|: \\ |\{x: book(x) \wedge by(douglas, x)\}|\end{matrix}\right) \geq d$$

The truth-conditions derived by this semantics can be paraphrased as: Mary has read many books by Douglas just in case the degree of books by Douglas that Mary has read is in the comparison class and is a superset of the degree of the neutral segment. Critically, the degree of the neutral segment depends on the contents of the comparison class in the context, as the difference in the elements of $[\![C]\!]$ in (15) and (16) illustrate. Thus, in Romero 2015b like Cohen 2001, both the regular and reverse interpretations of proportional *many* can be derived from one conservative denotation simply by assuming different Focus-marked patterns of emphasis. While Romero 2015b does address the cardinal interpretation of *many*, she posits a related but distinct denotation for it, given in (17):

(17) $[\![many_{card}]\!] = \lambda\, d_d. \lambda\, P_{\langle e,t \rangle}. \lambda\, Q_{\langle e,t \rangle}. (|P \cap Q|) \geq d$

One novel claim in Romero 2015b is that there is an analogous reverse interpretation for cardinal *many*, just as there is for proportional *many*, derived via the same Focus machinery as above.

Finally, Romero 2015b extends this degree-based approach to *most*, following on from work by Hackl 2009 (see also Solt 2015), claiming that *most* is simply the superlative form of *many*. In short, the composition and derivation of a *most*-utterance would proceed as it does for *many*, save for one critical difference: instead of *POS*, the degree morpheme is now -*est*, denoted as in (18).

(18) $[\![-est]\!] = \lambda Q_{\langle dt,t \rangle}.\lambda P_{\langle d,t \rangle}: P \in Q. \forall Q \in Q[Q \neq P \rightarrow Q \subset P]$

This denotation, when composed with a *many*-phrase like (15) would result in truth-conditions that can be paraphrased as: Mary has read most books by Douglas just in case the degree of books by Douglas read by Mary is a proper superset of every degree set of books read by other people (i.e. *Q* must be distinct from and larger than all the other degree sets in *P*). This account predicts that not only should *most* have analogous regular and reverse readings, it may also have a cardinal and proportional distinction, depending on which denotation of *many* is used. Romero 2015b uses cross-linguistic data to show that *most* in principle does have both regular and reverse readings but apparently only for the cardinal denotation of *many.* However, for syntactic reasons English does not allow Focus to associate with a phrase inside the DP where -*est* was base generated, thereby preventing the reverse reading from surfacing. Thus *most*, like *many*, is context sensitive, but for syntactic reasons *most* might not have exactly the same profile of interpretations as *many*.

As **Table 2-1** summarizes, while some theorists have posited distinct denotations to capture the different observed meanings of *many*, others have noted that pragmatic factors like Focus seem to disambiguate between meanings in principled ways. For example, Romero 2015b does somewhat reduce the proposed ambiguity of *many*, but she still assumes a lexical distinction between the proportional and cardinal interpretations of *many*. Without a clear description of how these two *many*s are related or derived from each other, this apparent ambiguity still poses a challenged to the theory of reference (see Parsons 1973). Therefore, after reviewing pragmatic machinery that may offer a more principled viewpoint on the role of Focus and information structure in deriving alternatives sets, the latter sections of this chapter propose a key innovation, adopting the degree-based approach in Romero 2015b that is able to derive the reverse proportional, regular cardinal, and reverse cardinal interpretations of *many*, all from one regular, relative, proportional, degree-based denotation of *many*.

Table 2-1. This table summarizes the accounts reviewed in **Section 2.3** and shows how each account captures the ambiguity and vagueness on *many*.

	Lexical ambiguity?	Vagueness?
Partee 1988	Yes – cardinal and proportional denotations	Yes, relevant comparison 'contextually defined'
Westerståhl 1985	Yes – regular and reverse denotations	?
Cohen 2001	No *	Yes, encoded as a relative comparison of proportions saturated by a comparison class defined by Focus-marking on the utterance
Romero 2015b	Yes – cardinal and proportional denotations	Yes, encoded as a relative comparison of proportional degree sets on a continuous, infinite scale; as in Cohen, the contents of these degree sets are saturated by a comparison class defined by Focus-marking on the utterance

* For the purposes of this dissertation; However, Cohen 2001 does discuss an absolute denotation, similar to those proposed by Partee 1988, where the proportion on the right side of the relation is an absolute value determined by context, e.g. what counts as *many* on average.

2.2. Pragmatic factors

As has been discussed above, *many* can take on several different interpretations, and at least some of these interpretations appear to depend on pragmatic factors like where emphasis appears in the utterance and what counts as relevant alternatives to the emphasized phrases in a particular context. While several of the semantic accounts in the previous section point to these pragmatic factors, they stop short of laying out formal mechanisms for capturing the behavior of these factors in a principled way. Below are reviewed several pieces of literature addressing in turn, a formal theory of discourse structure that then enables a description of what kinds of information can be contained in a discourse (**Section 2.2.1**), and how the structure and content of a discourse can specify relevant alternatives and license a particular interpretation of a *many*-utterance. Then **Section 2.2.2** discusses how information structure licenses phonological emphasis, and **Section 2.2.3** discusses the semantic effects of information structure, or Focus, introducing a semantic approach to the denotation of questions, which is a critical piece of the analysis in **Section 2.3** and the main theoretical innovation of this dissertation.

2.2.1. Formal accounts of discourse: Structures and Participants

This dissertation follows other researchers in making certain simplifying assumptions in order to formally describe discourses (c.f. Ginzburg 1997), namely: a turn consists of only one move (either one assertion or one question) – this model does not have the machinery to explain how a turn with more than one move can be accommodated into a participant's representation of the discourse; furthermore, perfect communication is achieved between discourse participants after every turn, i.e. the content of each turn is faithfully accommodated by all participants.

A discourse participant, intuitively, is anyone involved in a discourse. For simplicity, assume there are only ever two people involved in a dialogue, so that at any given time one is the hearer and one is the speaker. The D(iscourse)-trees shown below reflect an utterance of the answer to an immediately preceding question. According to Ginzburg 1997, each participant is endowed with the ability to represent key elements that together serve to define the formal structure of a discourse, depicted in **Figure 2-1**. The participant's GAMEBOARD includes elements about the discourse structure and where the participants sit in that structure, including what has been discussed previously in the discourse (FACTS, in **Figure 2-1**) what is currently under discussion (QUD in **Figure 2-1**), what was just uttered (LATEST-MOVE in **Figure 2-1**), and also a host of more general things pertaining to their internal UNPUBLICIZED MENTAL STATE (e.g. their own state of knowledge, their goals in the discourse, as well as potentially representations taken from other cognitive systems, like the ANS.).

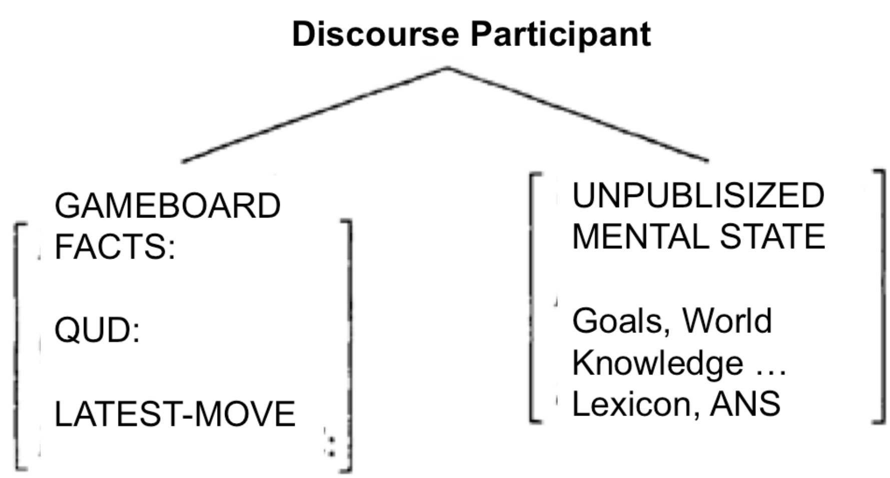


Figure 2-1. Representation of what kinds of information a discourse participant is able to represent. Each of these stores could be thought of as a kind of stack (à la Roberts 1996/2012, a.o.). Adapted from Ginzburg 1997.

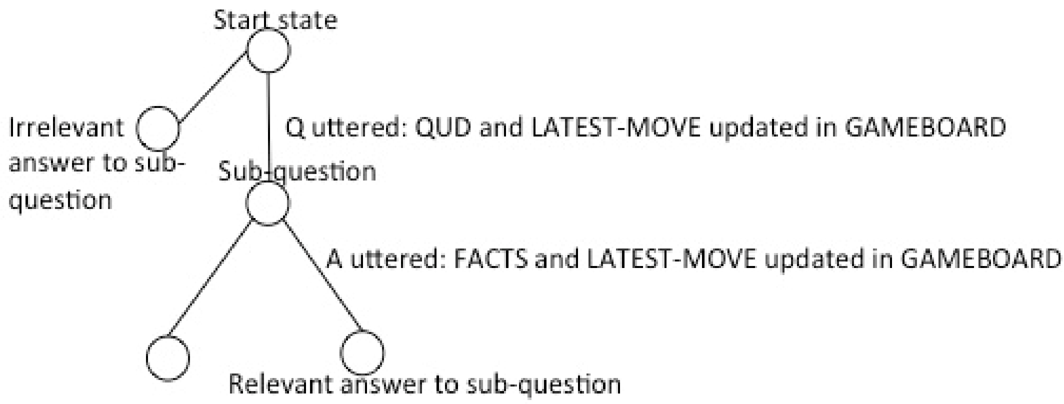


Figure 2-2 A generic D-tree as it relates to what a participant can represent.

Using the terminology above, we can define the structure of a discourse based on the information the participant is able to track, i.e. the discourse from their perspective. To illustrate this structure, this proposal adopts the formal representation of D-trees (c.f. Büring 2003; Krifka 2001; see also Cohen & Krifka 2014). Referring to **Figure 2-2**, discourses can be thought of as tree-like structures, with the highest node representing the over-arching goal or the initiating question of the discourse, and each line emanating from it representing a possible move (a sub-question or an answer) that would narrow the range of possible answers to that larger question. In this way, each node – a 'state' in the discourse – can be thought of as reflecting what has been uttered before in the discourse, encoded by the elements of the participant's GAMEBOARD and MENTAL STATE as described above. As another turn is taken, the contents of that turn (i.e. the FACTS asserted, a QUD raised, etc.) are accommodated into the participants GAMEBAORD, and the participant 'moves' to the next state in the discourse. Every turn links a question-answer pair, but the question may well be implicit (Stalnaker 1978). Moreover, implicit moves may be accommodated into previously built parts of the D-tree in order to support, e.g. inferences about a participant's intended QUD. In short, this accommodation process is a formal means of representing how speakers coordinate their D-trees when, for example, things like Gricean quantity implicatures are calculated. Thus, the content of each state and each turn naturally can be thought of as coming from the linguistic (or otherwise) information uttered (or implied), as discussed further below.

This formalism also provides the machinery necessary to define the felicity conditions of question-answer pairs. For example, relevance can be defined via the structure of the D-tree. If we accept the stipulation that only downward moves are allowed, we can adapt Roberts' 1996/2012 definition of relevance to a D-tree structure as follows: only nodes immediately dominated by the current state are relevant to the QUD of that state, because they are the possible answers to that QUD and the QUDs of the states dominating it.

Building off of this notion, relevant questions can also be resolving questions. As Ginzburg 1997 observes, resolvedness actually emerges as a presupposition derived from the goal(s) of the participant, e.g. "I wondered about *q*, so I asked *q*. She told me that … and …etc. This was true. The question *q* is now resolved, so now I know *q*." (p. 25). Thus, resolving answers are relevant answers in the D-tree that lead to the participant knowing some fact that they previously wondered about. Implicit in Ginzburg's discussion of this issue is the idea that to wonder about something somehow reflects the participant's goals, but Ginzburg does not elaborate much on what goals can be, and exactly how goals are manifested in the participant's representation of the discourse so that they are able to give rise to such resolvedness presuppositions.

van Rooy (2003) proposes that this context dependency of resolvedness can be modeled by defining an expected utility function for actions after learning some fact and assuming rational agents want to maximize utility. He then defines a strict order over information that then orders all relevant answers to a question in terms of utility in the context of a particular decision problem (which he assumes is defined by the Hamblin-style partition semantics of questions, Hamblin 1973). Thus, in effect, questions set up decision problems that have solutions in the form of answers ranked according to the utility of the information they provide relative to the decision problem.

So, goals can be thought of as specifying a particular utility function that ranks possible answers. In this way, they play a critical role in the account developed in **Section 2.3**. On the one hand, goals determine the appropriate granularity of an answer. That is, they license either a *many*-utterance or an exact cardinality answer to a "How many" question. Additionally, this dissertation posits that goals – in coordination with the QUD and the participants' world knowledge – order the possible answers denoted by a "How many" question, essentially delimiting what counts as relevant alternatives to the quantified phrase; the comparison class. For example, if a goal specifies that the speaker wants information about the three most common, large land-dwelling mammals that live in China, rats will be ranked very low in the scale of resolving answers, because they are not large for land-dwelling mammals. In short, goals provide a critical tool for defining why *many*-utterances can be relevant, resolving answers to "How many" questions and why particular potential answers are not considered relevant alternatives, and therefore should not enter into the derivation of meaning for *many*-utterances.

In sum, elements of the participants' MENTAL STATE and the D-tree affect whether a *many*-utterance is felicitous (or whether a more precise answer is sought), and how information relevant to the semantic denotation of *many* may be represented. The next sections discuss the phonological and semantic effects of licensing QUDs, and how the values of the variables in the semantic denotation of *many* are selected from the information structure of the preceding discourse, defined by relevant, resolving answers. The way these values enter into the composition of the truth-conditions of a *many*-utterance is also discussed. Then, the final section considers how these mechanisms work in tandem to explain how the various phenomena observed in **Section 2.1** arise.

2.2.2. Formal accounts of Focus and Topic: The phonological signaling of information structure

Before turning to the semantic denotations of licensing QUDs and *many*, some literature on Focus and Topic is reviewed, in order to more clearly define what these terms mean and explicate their role in this account. The broad consensus in the literature is that the notion of Focus (and its counterpart, Topic) are closely linked to the information structure of the preceding discourse (Schwarzschild 1999; Wagner 2012; a.o.). In various theories, Focus can be marked with a particular prosodic contour, a syntactic location, or a particle (Büring 2003; Choi 1996; Hendriks 2004; Herburger 2000; Tomaszewicz 2016; Wagner 2012); in some theories these markers effect structural changes at LF (Choi 1996; Tomaszewicz 2016; Wagner 2012), in others they introduce new material into the semantic composition (e.g. von Fintel 1994; Krifka 1995; Rooth 1992). Some accounts of the effects of Focus on interpretation treat these issues as one in the same (Büring 2003; Rooth 1992); others separate the prosodic from the semanto-syntactic (Hendriks 2004; Szabolcsi 1986; Tomaszewicz 2016). The account developed here assumes a close relationship between prosody and information structure. However, none of the major implications of this account are compromised if this relationship turns out to be more tenuous than described here.

According to Schwarzschild 1999 all pragmatically new information in the discourse must be Focus-marked, but not all Focus-marked constituents represent new information. To formally achieve this result Schwarzschild proposes two competing, violable constraints, and shows that they operate in tandem, with GIVENNESS outranking AVOIDF:

GIVENNESS: If a constituent is not F[ocus]-marked, it must be GIVEN.

AVOIDF: F[ocus]-mark as little as possible, without violating GIVENNESS.

Therefore, Schwarzschild 1999 concludes that Focus-marking can be indicative of pragmatically new information, but sometimes appears on old information in the absence of any 'newer' information to mark. This conclusion suggests that Focus is one mechanism by which the congruity of question-answer pairs in a discourse can be assessed, and particular prosodic contours[8] (e.g. Büring's 2003 A- or B-accents) belie this underlying relation to information structure in the discourse.

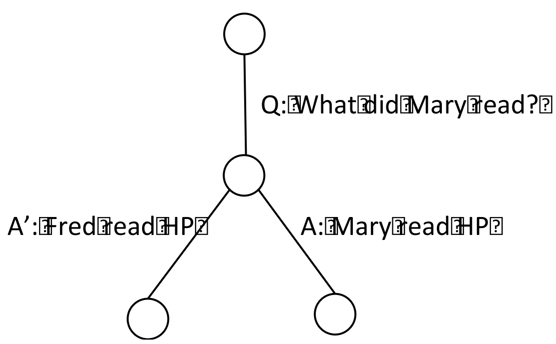
Büring 2003 shows that the B-accent pattern associated with a Contrastive Topic (CT) indicates a particular 'strategy' in the discourse (Jackendoff 1972) – more

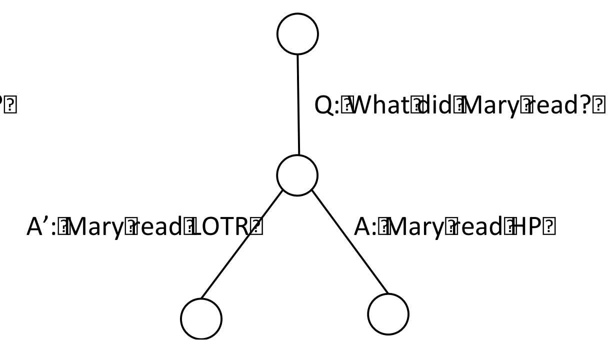
[8] In English; in other languages, Focus can be realized with particles (e.g. Japanese; Kuno 1973) or scrambling (German & Korean: Choi 1996; Hungarian: Kiss 2008). These facts provide hints about the syntactic behavior of Focus at LF and how it interacts with other scope-taking elements as discussed further below and in the **Discussion** of **Chapter 3**.

specifically, CT-marking indicates the form of the question (implicit or explicit) that must precede the utterance in the D-tree (Stalnaker 1978), henceforth the licensing QUD. Formally, Büring defines a function $[\ ]^{CT}$ that when applied to a declarative utterance with CT-marking generates a set of alternative questions (the 'CT-value') that can be said to license that utterance (see also the alternative semantics of Rooth 1992). For example, the different CT-markings on (19a) and (19b) below belie differently structured D-trees, indicating that different strategies are being pursued in the two different discourses (but both discourses are answering the same over-arching question, "Who read what?").

(19) a. Q: What did Mary read? b. Q: What did Mary read?

A: MARY_{CT} read Harry Potter. A: Mary read $[\text{HARRY POTTER}]_{CT}$





CT-value of (19a) A: {What did Mary read, What did Fred read…}

CT-value of (19b) A: {Did Mary read Harry Potter, Did Mary read Lord of the Rings…}

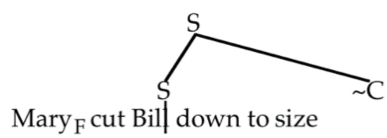
That is, the question can be thought of as 'generating' a particular structure of D-tree, and the CT-marking picks out some sub-tree in the D-tree. So again, overt phonological stress marking is licensed by, or is anaphoric to (c.f. Rooth 1992, von Fintel 1994), elements in the information structure of the discourse.

According to von Fintel 1994, the context sensitivity can be equated to the Focus-marking in an answer, which is anaphoric to elements in the preceding discourse context. According to von Fintel 1994, 2004 and Rooth 1992, the Focus operator ~ takes two arguments, the constituent or element marked with the B-accent and a proposition and produces a set of alternatives to that proposition. For example, in (20), the Focus-marking on Mary combined with the presence of the ~ operator at LF produces the anaphoric variable C that is now looking for an antecedent of the form ‘x cut Bill down to size:

(20) Q: Who cut Mary down to size?

A: $[\text{Mary}]_F$ cut Bill down to size.

A at LF:

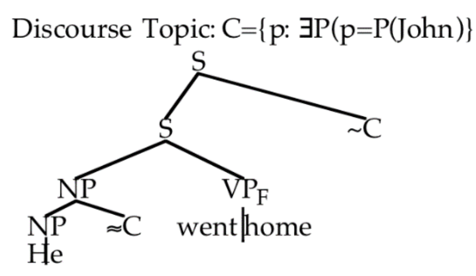


Similarly, von Fintel’s Topic operator ≈ again takes the same two arguments and now the anaphoric variable C is looking for an antecedent that is a subset of the set of propositions of the form ‘Φ Ps’ (e.g., ‘John Ps’ in (21)).

(21) Q: What did John do?

A: $[\text{He}]_T$ $[\text{went home}]_F$.

A at LF:

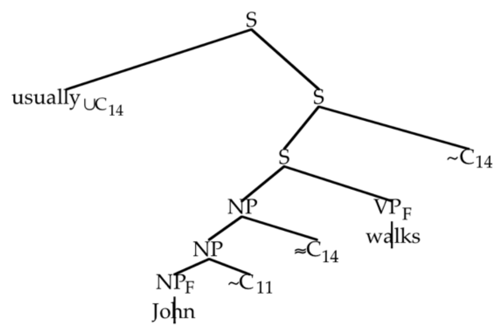


Therefore, Contrastive Topic of the kind discussed by Büring 2003 occurs when Focus occurs under Topic in the LF of an utterance, as in (22), now there is an anaphor looking for a set of subsets of the set of propositions of the form 'John gets to school in manner x.':

(22) Q: How do students around here get to school?

A: $[[\text{John}]_F]_T$ usually $[\text{walks}]_F$

A at LF:



Where C_{14} = {p: ∃x (p= John gets to school in manner x.)}

Note that this amounts to the same set of propositions derived by Büring's 2003 $[\ ]^{CT}$ function, .

In the account developed below,

To summarize, phonological realizations of emphasis belie the underlying representation of the speaker's information structures in a discourse: what FACTS they think they are introducing as new to the discourse, and what accommodated FACTS (from the comparison class generated by the licensing QUD) they wish to contrast or re-raise. As we will see below, these are important constructs for the interpretation of *many*-utterances, but as von Fintel 1994 and 2004 argue the apparent link between pragmatic concepts like information structure and quantifiers is not the result of the grammaticalized encoding of pragmatic information in the semantic denotation of quantifiers, but is instead the result of coincidental anaphoric reference to the same types of information in the discourse, via the ~ operator at LF. For example, he proposes that Focus-marked constituents are anaphoric to the same discourse information as quantifier domains (the sister of the quantifier in its base-generated position). von Fintel 1994 shows that this apparent correspondence between Focus-marked elements and the quantifier domains is a mere coincidence, licensed by the underlying pragmatic information structure of the discourse, but that no reference to Focus should be necessary in the semantic machinery of quantifiers.

The analysis in **Section 2.3** continues in this vein, positing that the prosodic B-accent is one way of signaling the presence of Focus-sensitive interpretations, in a spoken discourse (Büring 2003). The Focus-sensitive reading is engendered by the ~ operator at LF that introduces contents from the information structure in the discourse context into the composition (von Fintel 1994; 2004); its presence and effects are detectable just when the information structure gives rise to new or contrastive information (Krifka 1995; Schwarzschild 1999). However, accounts such as von Fintel's and Rooth's all face questions about how the contents from the discourse in fact compose with the rest of the utterance in the meta-language (e.g. lambda-functions). The following section introduces the notion of structured meanings to solve this problem and offers a transparent link between licensing QUDs and Focus-marking. Then the following section describes the novel account of the regular-reverse ambiguity proposed in this dissertation.

2.2.3. The Structured Meaning Denotation of a question: A formal semantic account of licensing QUDs of Focus

As mentioned above, each state in a discourse can be reached by some kind of move, and the contents of that state can seemingly enter into the composition of the utterance, but in order for this pragmatic machinery to interface with the linguistic content of that move, the propositional content of each node must take a particular form. Many of the pragmatic approaches summarized above assume a propositional structure to the elements of the set denoting a question meaning (as in (19) above), following Hamblin 1973). However, the analysis developed in this dissertation seeks to develop a Krazterian-style denotation for use in a compositional derivation, i.e. lambda calculus functions. So, in order to interface with this denotation, the linguistic information of each turn must also be in the form of functions. To this end, this account adopts the structured meaning account of question denotations advocated in Krifka 2001 (see also Aloni, et al., 2002 for a similar approach). Under this account, the denotation of a question is a tuple comprised of the background function (a function that essentially defines the domain of the question) and the restrictor (a set of individuals, or the range of the function). An example is given in (23):

(23) Which books did Mary read? =

$\langle \lambda x . \text{Mary read } x, \text{BOOKS} \rangle$

Using the example (23), we can see that the meaning of a constituent question is simply a function, and the possible answers can be derived by applying each of the elements in the restrictor set, BOOKS, to the function. As an example, assume that the only elements of READ are *Harry Potter* and *Lord of the Rings*, then the possible answers would be {[Mary read Harry Potter.], [Mary read Lord of the Rings.]}. Note that these possible answers are propositionally analogous to the set of questions derived using the $[\ ]^{CT}$ function in Büring 2003, when applied to the utterance "Mary read Harry Potter$_{CT}$," as in (24).

(24) [Mary read Harry Potter$_{CT}$.]^{CT} =

$$\{\ \langle \lambda\, x\,.\, Mary\ read\ x, HP \rangle, \langle \lambda\, x\,.\, Mary\ read\ x, LOTR \rangle\ \}$$

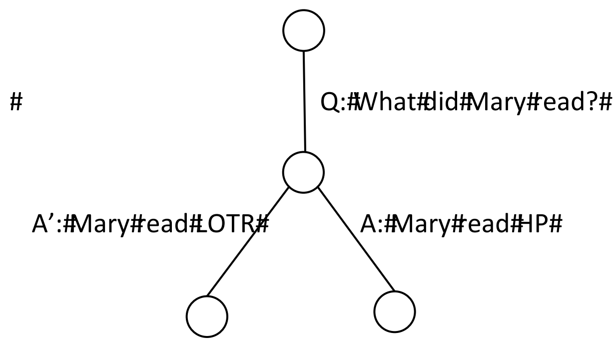


Figure 2-3. The D-tree for the discourse described in (19) and (24).

Büring 2003 does not speculate on how exactly this $[\ ]^{CT}$ function determines these values (Rooth 1992 treats them as presuppositions), but in the analysis proposed here the function in the structured meaning of the licensing QUD is defined relative to the goals of the speaker, and in turn generates the structure of the D-tree; the terminal nodes of the tree are populated by the members of the restrictor set: the relevant, resolving answers.

Furthermore, the account developed here follows Krifka 1995 (see also Aloni, et al., 2002), by assuming that the denotation of the Focus operator ~ could also be a structured meaning, making the link between licensing QUDs and the interpretation of Focus in Focus-sensitive utterances even more transparent and obvious. Krifka 1995 proposes that

Focus operators (e.g. covert ~ or *only*), partition the semantic representation of an utterance into a background and a focused element. Adapting Krifka's notation, the structured meaning of an utterance like (25) with the prosodic marker of Focus here signaled by bolded text would be as in (26).

(25) Mary read **Harry Potter.**

(26) $\sim(\langle \lambda\, x\, .\, \text{Mary read x}\rangle, BOOKS) =$

$\langle \lambda\, x\, .\, \text{Mary read x}\rangle\ (HP), where\ HP\ \in of\ BOOKS$

In other words, the Focus operator ~ partitions the elements of the utterance into a background function ($\lambda\, x\, .\, \text{Mary read x}$) and a Focus set, and when the background function is applied to the set, it derives a set of functions, one of which is the denotation of the utterance itself, with the Focus-marking serving to signal this set membership. Note that (26) is exactly the same structured meaning given by the licensing QUD of (25) in (23) above, suggesting that licensing QUDs and Focus are one in the same phenomenon.

The account developed below adopts structured meanings to denote the licensing QUDs that give rise to the particular information structure that in turn gives rise to particular interpretation of a *many*-utterance. This is the major theoretical innovation of the account developed in this dissertation. By virtue of representing the licensing QUDs in the preceding discourse as structured meanings, the phenomenon referred to as Focus, can be reduced to or thought of solely in terms of the information structure of the preceding discourse.[9]

To summarize, Focus-marking, for example Büring's B-accent in English, can signal new information in the discourse. Focus, implemented as a covert operator (~) at LF denoted as a structured meaning of the immediately preceding licensing QUD introduces a comparison class to the composition of a *many*-utterance, optionally resulting in the B-accent during pronunciation (Büring 2003; Krifka 1995; Rooth 1992). The comparison class is the set of relevant, resoling answers to the licensing QUD (Roberts 1996/2012;van Rooy2003; or congruent answers from Aloni, et al., 2002). Hence, Focus is licensed by a particular information structure in the preceding discourse context, because it is that information structure. In combining these formal tools, the analysis developed below provides formally compatible antecedents for the anaphoric ~ operator (i.e. propositions with unbound variables) and offers a formal explanation of the close relationship between discourse information structure and Focus, as well as how the various interpretations of *many*-utterances are derived from one denotation of *many*.

---

[9] Assuming that structured meanings can denote licensing QUDs and the semantic structure of Focus, suggests that the semantic and pragmatic effects of Focus may simply emerge from the composition of an utterance in the context of a particular discourse with a particular information structure. While this dissertation leaves further exploration of this issue for other researchers, structured meaning representations of Focus could also help disentangle issues of how Focus is marked (prosodically or syntactically) from the interpretive effects of Focus explored here (see Hendriks 2004, Hendriks & de Hoop 2001 for other suggestions on how to approach this issue). Since this is not the primary aim of this dissertation this issue is left aside, and the assumption is adopted that the prosodic B-accent is a reliable marker of the kind of Contrastive Topic Focus critical to the regular-reverse ambiguity investigated in this chapter; but considering the results of the experiments in **Chapter 3**, this may be a faulty assumption generally.

2.3. Putting it all together: A formal pragmatic-semantic account of *many*

This section outlines an account of how the regular and reverse proportional interpretations of *many* can be derived from the same underlying semantic denotation, then it extends this account to derive the cardinal interpretation from the same proportional denotation, thereby reducing the lexical ambiguity of *many* to purely pragmatic factors. This proposal draws on observations reviewed above (Cohen 2001; Herburger 1997; 2000; Romero 2015a; 2015b) that different patterns of prosodic emphasis, assumed to mark Focus, differentially give rise to distinct interpretations of *many*. Moreover, the account proposed here follows the accounts reviewed above in using Focus to derive multiple interpretations from one denotation. However, the account proposed here innovates on the previous accounts by showing that it is possible to derive all four of the proposed interpretations of *many* from one single denotation, by integrating the notion of Focus, how it relates to the information structure of a discourse (critically including the participants goals and internal knowledge states), and the licensing QUD in a discourse.

The analysis below assumes the denotation of *many* and the silent degree morpheme *POS* in (27) and (28), inspired by Romero 2015b, but with minor adjustments motivated by a fully articulated semantic composition (see **Figure 2-6**). Specifically, the order of the arguments in $[\![many_{prop}]\!]$ was altered so that the degree argument remains unsaturated until its composition with $[\![POS\ C]\!]$ since one major claim here is that the value of that degree variable comes from the comparison class derived from the discourse information structure, introduced by ~ at LF.

(27) $[\![many_{prop}]\!] = \lambda P_{\langle e,t\rangle}.\lambda Q_{\langle e,t\rangle}.\lambda d_d.(|P \cap Q|:|P|) \geq d$

(28) $[\![POS]\!] = \lambda Q_{\langle\langle d,t\rangle,t\rangle}.\lambda P_{\langle d,t\rangle}.P \in Q.L_{\langle\langle dt,t\rangle\langle d,t\rangle\rangle}(Q) \subseteq P$

For illustration, the following analysis will make reference to the data points in (29), where emphasis is marked by boldface and subscripted F/CT, the truth-conditions that arise based on this Focus-marking are paraphrased below the sentence (adapted from Herburger 2000, Partee 1988, Romero 2015b, and Westerståhl 1985), and the name for the interpretation is provided in italics:

(29) Many Scandinavians have won the Nobel Prize.

c. Many Scandinavians have won the **Nobel Prize** $_{F/CT}$. *regular proportional*

Paraphrase: Of all Scandinavians, the proportion that have won the Nobel Prize in Literature is larger than the proportion of them that have won other things.

d. Many **Scandinavians** $_{F/CT}$ have won the Nobel Prize. *reverse proportional*

Paraphrase: Of all the people that have won the Nobel Prize, the proportion of them that have been Scandinavians is larger than the proportion of winners from other countries.

Recall that a D-tree is a representation of a discourse, representing how questions have been raised and answered (Büring, 2003, Cohen & Krifka, 2014; Krifka, 2001). Participants each have internal mental states, which represent – among other things – goals that specify the level of precision required for an answer to be resolving (c.f.

Ginzburg, 1997; Roberts 1996/2012; van Rooy, 2003). This analysis posits that within these D-trees, *many*-utterances are taken to be licensed by – or anaphoric to – dominating questions of the form "How many…." Thus *many*-utterances are relevant (Roberts 1996/2012), resolving (van Rooy, 2003) answers to "How many…" questions; henceforth referred to as licensing QUDs. One critical innovation adopted by this analysis is to denote licensing QUDs as structured meanings following Krifka 2001, so that for example the licensing QUD in (30) is denoted as (31):

(30) How many Scandinavians have won different honors?

(31) $\langle \lambda\, d\, .\, (\{|x: won(x, Scandinavians)|) \geq d\}, DEGREES\rangle$

The participant's world knowledge provides the known alternatives, represented as separate sub-trees in the D-tree (**Figure 2-4**), that are derived by applying the background function to the restrictor set of this structured meaning, giving rise to a set of degree functions. These functions represent the set of relevant answers to the licensing QUD, each represented by one of the terminal nodes in the D-tree. This set of degree sets derived from (31) is shown in (32) and the D-tree that licenses it is illustrated in **Figure 2-4**:

(32) $[\![C]\!] \subseteq \{\lambda d'.\,(|\{x: Scandinavians(x)\} \cap \{x: win(NP, x\}|: |\{x: Scandinavians(x)\}|) \geq d,$
$\lambda d'.\,(|\{x: Scandinavians(x)\} \cap \{x: win(goldmedal, x\}|: |\{x: Scandinavians(x)\}|) \geq d \ldots\}$

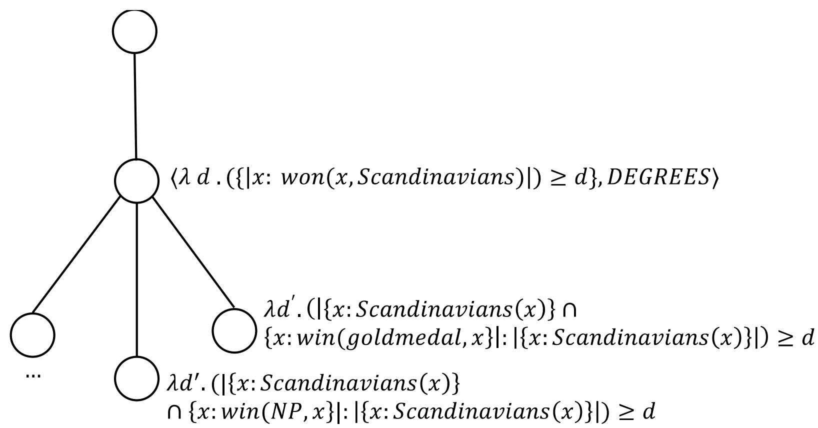


Figure 2-4. The D-tree that licenses the regular interpretation, with alternative degree sets (comparison class) licensed by (30) and derived from (31).

Thus, this set of alternatives (the terminal nodes), referred to as the comparison class (c.f. Rooth 1992), that is defined by the licensing QUD is the referent of the anaphoric Focus operator, and it is in this sense that the licensing QUD licenses one or another Focus-marking pattern. That is, the structure of the D-tree – the number and contents of the terminal nodes – constrain what Focus can associate with. In short, the licensing QUD, by virtue of the comparison class and D-tree it gives rise to, defines the felicity conditions of *many*. To see this, consider the situation in (33).

(33) A: How many Scandinavians have won different honors?

B: # Many **Scandinavians** $_{\mathbf{F/CT}}$ have won the Nobel Prize.

This account predicts that B's answer is infelicitous given A's question, which seems intuitively correct. This infelicity arises because B's utterance contains Focus-marking on a phrase for which no alternative degree sets exist in the D-tree; there are only sets that describe the degree to which Scandinavians have won various honors – there is nothing to contrast Scandinavians with. Of course, this discourse can be saved. An alternative D-tree can quickly be constructed by a cooperative participant. This would require A to reconstruct or infer the licensing QUD that licenses B's utterance, i.e. the licensing QUD for the reverse interpretation. This reconstructed licensing QUD and its denotation are given in (34) and (35):

(34) How many people of each nationality have won the Nobel Prize?

(35) $\langle \lambda\, d\, .\, (\{|x{:}\, win\ (NP', x)|) \geq d\}, DEGREES\rangle$

This newly constructed comparison class and D-tree are shown in (36) and **Figure 2-5,** respectively:

(36) $[\![C]\!] \subseteq$

$$\{\lambda d'.\,(|\{x{:}\, Scandinavians(x)\} \cap \{x{:}\, win(NP, x\}|{:}\, |\{x{:}\, Scandinavians(x)\}|) \geq d,$$

$$\lambda d'.\,(|\{x{:}\, Andorrans(x)\} \cap \{x{:}\, win(NP, x\}|{:}\, |\{x{:}\, Andorrans(x)\}|) \geq d \ldots\}$$

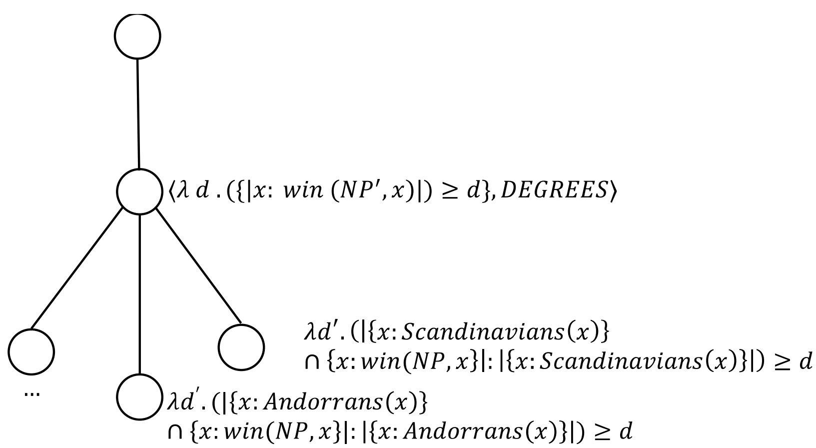


Figure 2-5. The D-tree that licenses the reverse interpretation, with the comparison class licensed by (34) and derived from (35).

In short, B was answering the question in (34) instead of A's initial question in (30). Under this account, B would be assumed to have world knowledge that only enabled them to generate the comparison class (36), which gives rise to a D-tree as in **Figure 2-5**, and therefore B's response would also give rise to a classic Gricean informativity implicature about B's knowledge state. The main point, however, is that different truth-conditions arise because of the different contents of the D-trees, derived from distinct licensing QUDs, and if these truth-conditions are not relevant, resolving answers to a licensing "How many…" question, then infelicity – or worse false-hood – arises.

That distinct truth-conditions can be derived from the same semantic denotation, simply by assuming different D-trees, can be seen by tracing the complete composition of the *many*-phrase with *POS*, as shown as in **Figure 2-6**.

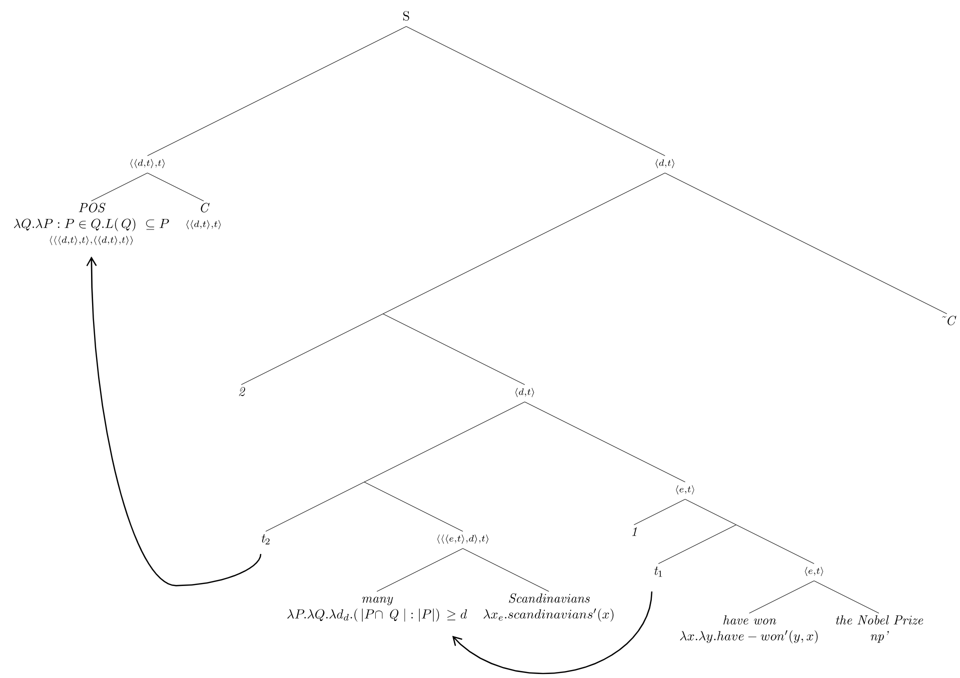


Figure 2-6. The LF composition of *Many Scandinavians have won the Nobel Prizes*. Note that in this LF, unlike those posited by Romero 2015b (see **Section 2.1**), *many* undergoes QR – one consequence of assuming that *many* is indeed a generalized quantifier-typed determiner which she does not explore.

Following the denotation of *POS* in Romero 2015b, this account assumes that *POS* contains a function *L* that operates over *Q* (the value of which is anaphorically bound to the set of degree sets in the context $[\![C]\!]$, via ~), and returns what Romero calls the neutral segment[10], which can intuitively be thought of as a measure of central tendency over the full comparison class. This formulation of *POS* means that a *many*-utterance evaluates to true just in case *P* is a superset of the degree set returned by *L(Q)*, i.e. *P* must meet or exceed upper bound the neutral segment. These truth-values are visually illustrated on the number lines below each derivation, which assume degrees in the set of rational numbers that would make each utterance true. First, consider the derivation of the truth-conditions for the **regular interpretation** in (2):

(37) Many Scandinavians have won the **Nobel Prize** $_{F/CT}$.

LF: [[POS C]] [1[t_2-many Scandinavians] [2[t_1] have won $NP_{F/CT}$]]] ~C]

$[\![C]\!] \subseteq$

$\{\lambda d'.(|\{x: Scandinavians(x)\} \cap \{x: win(NP, x\}|: |\{x: Scandinavians(x)\}|)$

$\geq d,$

$\lambda d'.(|\{x: Scandinavians(x)\}$

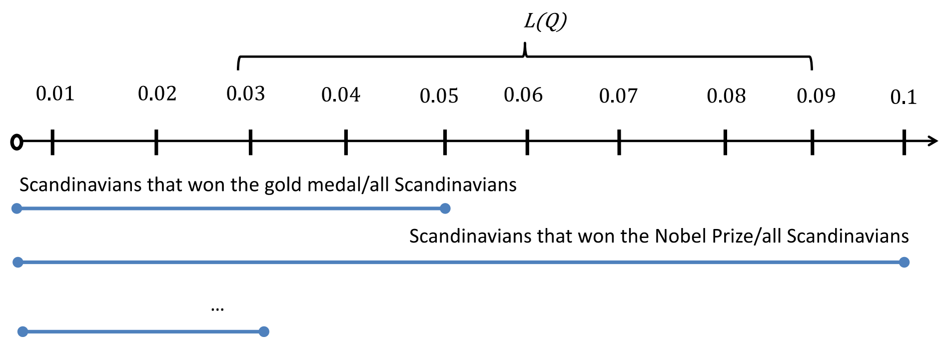
$\cap \{x: win(goldmedal, x\}|: |\{x: Scandinavians(x)\}|) \geq d \ldots\}$

$L([\![C]\!]) \subseteq$

$\lambda d'.(|\{x: Scandinavians(x)\} \cap \{x: win(NP, x\}|: |\{x: Scandinavians(x)\}|)$

$\geq d$

[10] Solt 2009 specifically addresses the correct way to conceptualize this neutral segment, noting that it must be capable of representing fine-grained, point-wise differences in distributions (e.g. it must convey both the central tendency and the variance). A more subtle consideration of this issue is left aside and this proposal assumes the neutral segment is described by the mean and standard deviation of the sets in the comparison class.



Next, consider the derivation of the truth-conditions of the **reverse interpretation** in (38):

(38) Many **Scandinavians** $_{F/CT}$ have won the Nobel Prize.

LF: [[POS C]] [1[t_2-many Scandinavians $_{F/CT}$] [2[t_1] have won NP]]] ~C]

$[\![C]\!] \subseteq$

$\{\lambda d'.(|\{x: Scandinavians(x)\} \cap \{x: win(NP, x\}|: |\{x: Scandinavians(x)\}|)$

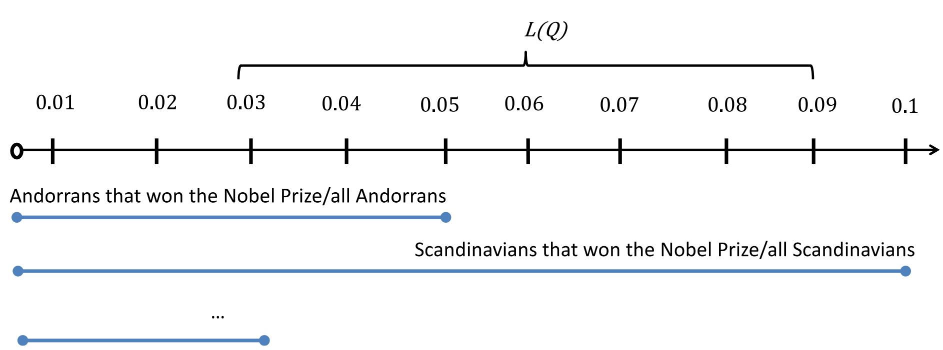
$\geq d,$

$\lambda d'.(|\{x: Andorrans(x)\} \cap \{x: win(NP, x\}|: |\{x: Andorrans(x)\}|) \geq d \ldots\}$

c. $L([\![C]\!]) \subseteq$

$\lambda d'.(|\{x: Scandinavians(x)\} \cap \{x: win(NP, x\}|: |\{x: Scandinavians(x)\}|)$

$\geq d$



These derivations show how simply changing the licensing QUD (which changes the D-tree), in turn changes the contents of the comparison class. The contents of the comparison class provide different degree sets to the semantic composition and thereby derive distinct truth-conditions using the same sematic denotation.

Given this fully developed account of the proportional interpretations of *many*, this section closes with a novel proposal that shows how the cardinal – reverse and regular – interpretations can similarly be derived using the same pragmatic-semantic machinery, instead of positing a completely different denotation. This account of *many* is also tentatively extended to *most* in the section that follows.

2.3.1. Extension to the cardinal interpretation

In order to derive the cardinal interpretation within this framework, recall that the contents of the comparison class are constrained not only by the licensing QUD but also by the participant's world knowledge and discourse-specific goals. Imagine an ignorant participant who knows nothing about what kinds of people win the Nobel Prize or anything about Scandinavians. Once the utterance in (39) is uttered to this ignorant participant, the derivation of the truth-conditions would proceed as in (39) and their D-tree will look like **Figure 2-7**.

(39) Many Scandinavians $_{F/CT}$ have won the Nobel Prize.

LF: [[POS C]] [1[t_2-many Scandinavians $_{F/CT}$] [2[t_1] have won NP]]] ~C]

$[\![C]\!] \subseteq$

$$\{\lambda d'. (|\{x: Scandinavians(x)\} \cap \{x: win(NP, x\}|: |\{x: Scandinavians(x)\}|) \geq d\}$$

c. $L(\llbracket C \rrbracket) \subseteq$

$\lambda d'.(|\{x: Scandinavians(x)\} \cap \{x: win(NP, x\}|: |\{x: Scandinavians(x)\}|) \geq$

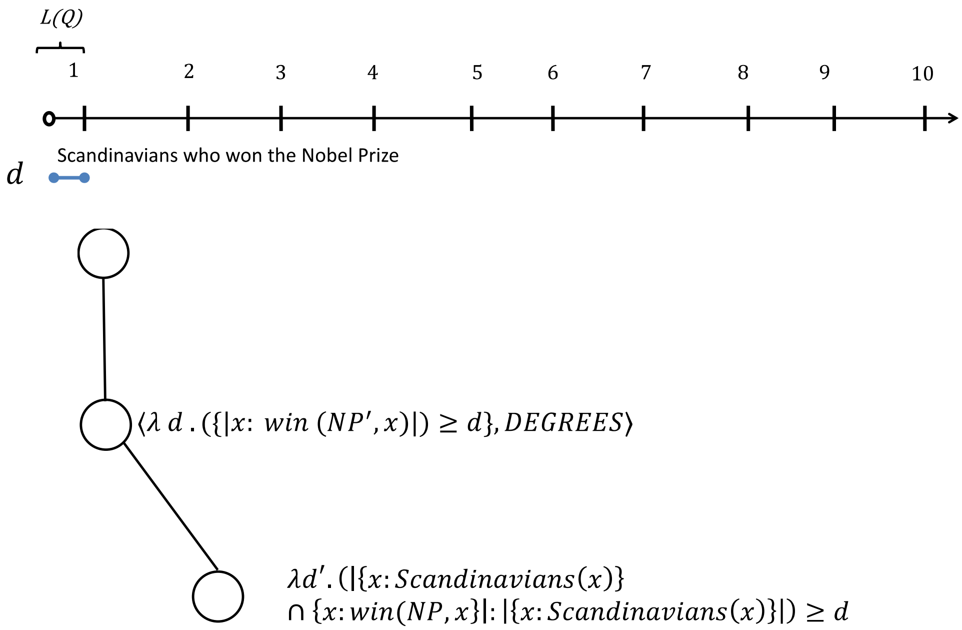


Figure 2-7. The D-tree of an ignorant participant after the question *How many people of each nationality have won the Nobel Prize?* is asked.

Note that the degree scale is now depicted as whole numbers – the critical difference here. The degree scale is being calculated in a different way because the participant has no knowledge about Scandinavians; they have no way of representing what proportion of all Scandinavians they are dealing with here. This in turn affects the way the neutral segment is calculated. Following Aloni, et al. (2002; see also Rooth 1992; Tomaszewicz 2016 a.m.o.) this proposal assumes that the licensing QUD for (39) presupposes $\lambda d .\{|x: win\ (NP', x)| \geq d\} \neq \emptyset$, so the participant can incorporate that into their D-tree, but the degree value of the set $\lambda d.\{|\{x: Scandinavians(x)\}| \geq d\}$ is still unknown. This proposal also assumes that the utterance of (39) presupposes that

$\lambda d.\{|\{x: Scandinavians(x)\}| \geq d\} \neq \emptyset$. This is illustrated as a single terminal node in the D-tree. Critically, given these presuppositions, this D-tree and singleton comparison class, the proportional meaning of $L([\![C]\!]) \subseteq \lambda d'.(|\{x: Scandinavians(x)\} \cap \{x: win(NP, x\}|: |\{x: Scandinavians(x)\}|) \geq d$ is just 1/1. Moreover, this utterance would evaluate to true, because the function L has only the one degree set to operate over, and so the neutral segment and the degree set have the same range, and therefore $L\ (Q) \subseteq P$ holds.

Now, imagine that we told this ignorant participant that three Scandinavians and two Andorrans had won the Nobel Prize. Critically, though, we did not tell them anything about how many total Scandinavians or Andorrans there are in the world. This means that the denominator in the proportion of each degree set is still just 1, which means that the degree scale is still being calculated in whole numbers rather than proportions. The resulting derivation and D-tree after this small amount of information is told to this ignorant participant is shown in (40) and **Figure 2-8** respectively:

(40) Many **Scandinavians** $_{F/CT}$ have won the Nobel Prize.

LF: [[POS C]] [1[t_2-many Scandinavians $_{F/CT}$] [2[t_1] have won NP]]] ~C]

$[\![C]\!] \subseteq$

$\{\lambda d'.(|\{x: Scandinavians(x)\} \cap \{x: win(NP, x\}|: |\{x: Scandinavians(x)\}|)$

$\geq d,$

$\lambda d'.(|\{x: Andorrans(x)\} \cap \{x: win(NP, x\}|: |\{x: Andorrans(x)\}|) \geq d\}$

$L([\![C]\!]) \subseteq$

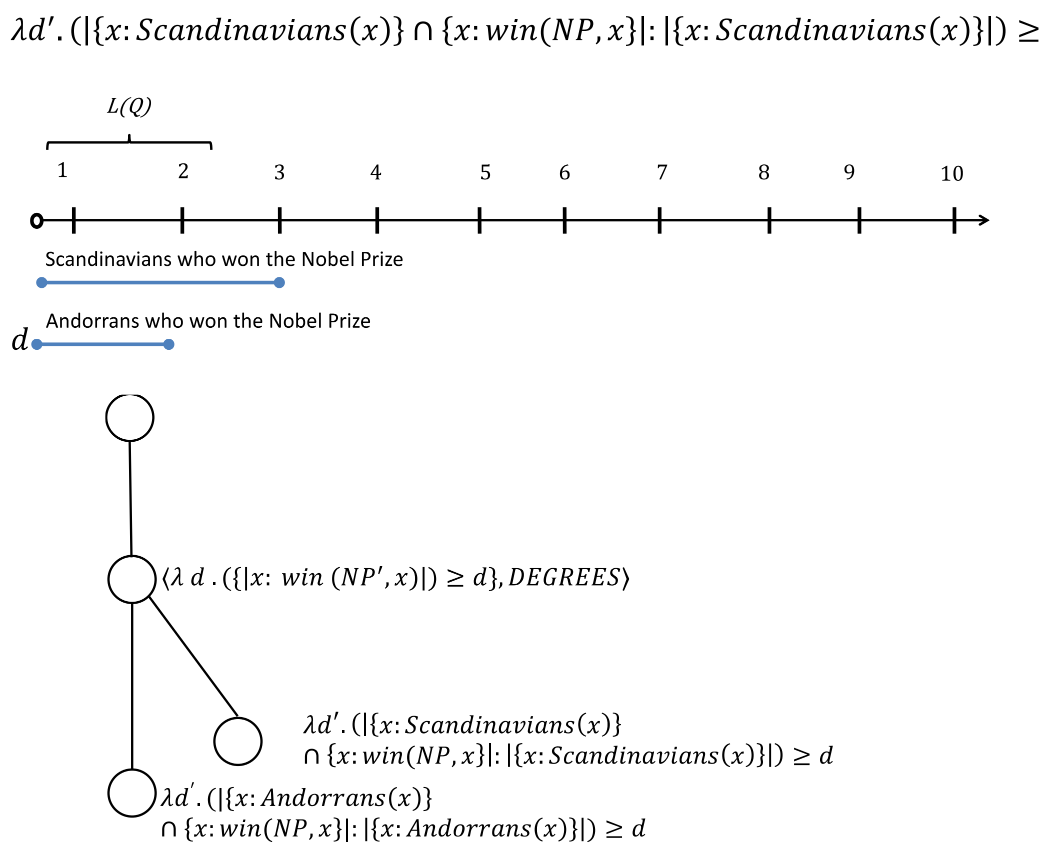
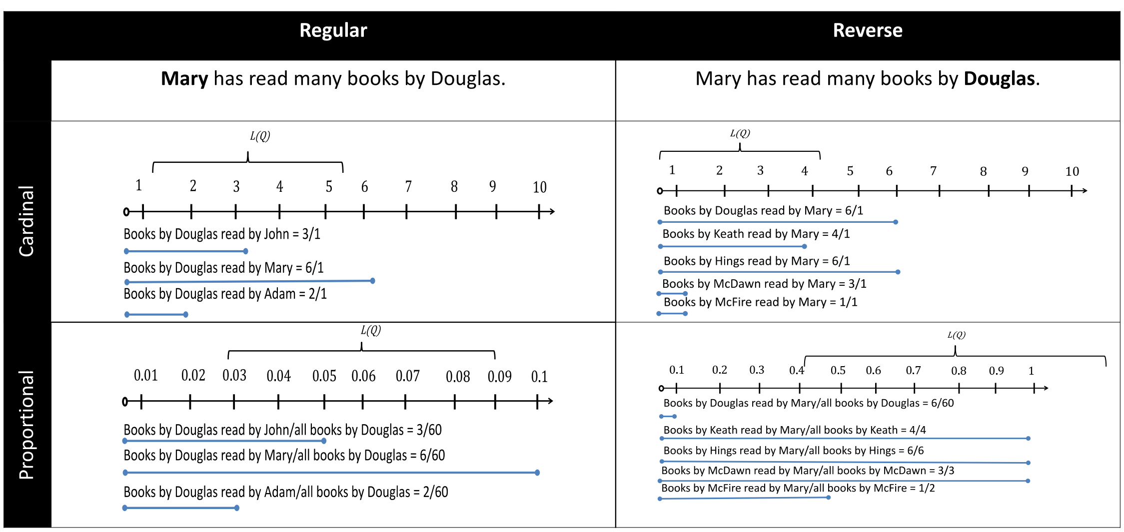


Figure 2-8. The tree after the ignorant participant has been told that three Scandinavians and two Andorrans have won the Nobel Prize.

In short, only once this ignorant participant learned something about the populations of Scandinavia and Andorra would their interpretation become proportional. To summarize, the cardinal interpretation – both the reverse and regular – arise when participants lack the knowledge that would specify the elements of the degree sets in the denominator of the proportional denotation of *many*. Therefore, we do not need to posit a distinct denotation, we only need to carefully consider the discourse context and knowledge state of the participants, in order to principally derive all four observed interpretations of *many*.

The account developed here draws on previous literature that strongly motivates a deterministic role for information structure in Focus-marking, and the role of information structure in pragmatic disambiguation generally. Using structured meanings to denote licensing QUDs in the discourse context, the account developed here posits that association with Focus, implemented via the ~ operator at LF, brings the contents of the comparison class into the composition, in order for it to compose with the degree morpheme *POS*. Thus, the information structure denoted by ~ in the composition gives rise to and disambiguates between the **regular** and **reverse interpretations** of *many*-utterances without needing to posit a distinct denotation for the quantifier. As **Table 2-2** summarizes, carefully characterizing the structure of the licensing QUD, and the comparison class it generates, further allows the derivation of a **cardinal interpretation** using an inherently proportional denotation of *many*, simply by accurately modeling the discourse information structure of an ignorant participant.

Table 2-2. This table summarizes the derivation of the four interpretations of *many* that have been observed in the literature, using the degree-based context sensitive approach developed in this **Section 2.3**.



2.3.2. Preliminary extensions to *most*

To evaluate the generality of the account developed in this chapter, this section compares and contrasts the predictions of previous accounts of *most* to an account suggested by the machinery proposed above, and explores what, if any, additional explanatory power the current account has over those previous. Ultimately, this discussion points to certain improvements to the account developed above that would be required in order to capture the observed data for *most*, but leaves a full implementation of these to future work.

Most does not appear to be sensitive to Focus in the same way *many* is, i.e. there is no reverse interpretation of (41) (i.e. where (41) would mean that 'Nobel Prize winners are Scandinavians for the most part'). Rather (41) can only convey that 'the majority of Scandinavians have won the Nobel prize' (Romero 2015b).

(41) Most Scandinavians have won the Nobel Prize.

Therefore, the question arises as to why the reverse interpretation is blocked, if we assume, like Romero 2015b, Hackle 2009, and others, that *most* is built from a denotation of *many*, and that *many* is Focus-sensitive. Romero 2015b claims that *most* is focus-sensitive, citing data from Pancheva and Tomaszewicz 2012 and Tomaszewicz 2013 from Bulgarian and Polish to demonstrate that the reverse interpretation does in fact surface, with a cardinal interpretation of *most*. However, it seems that in English the Focus mechanism is blocked by some element that is not present in these Slavic languages.

First, a sketch of Romero's 2015b account is provided, then the differences between Romero's account and the current one are explored, before considering limitations of the current account revealed by *most*. Recall from **Section 2.1** that Romero 2015b also extends her degree-based approach to *most*, following on from work by Hackl 2009 (see also Solt 2015), claiming that *most* is the result of composing $[\![many]\!]$ with the superlative morpheme $[\![-est]\!]$ (the denotation repeated here as (42) for convenience) instead of the positive morpheme $[\![POS]\!]$.

(42) $[\![-est]\!] = \lambda Q_{\langle dt,t\rangle}.\lambda P_{\langle d,t\rangle}: P \in Q. \forall Q \in Q[Q \neq P \rightarrow Q \subset P]$

Assuming this denotation, and that of *many*, repeated here as (43), the derivation of the truth-conditions for a *most*-utterance with Focus on *Nobel Prize* (i.e. the regular interpretation pattern) would proceed as in (44).

(43) $[\![many_{prop}]\!] = \lambda P_{\langle e,t\rangle}.\lambda Q_{\langle e,t\rangle}.\lambda d_d.(|P \cap Q|:|P|) \geq d$

(44) Most Scandinavians have won the **Nobel Prize** $_{\textbf{F/CT}}$.

a. LF: [[POS C]] [1[t_2-many Scandinavians] [2[t_1] have won $NP_{F/CT}$]]] ~C]

b. $[\![C]\!] \subseteq$

$\{\lambda d'.(|\{x: Scandinavians(x)\}$

$\cap \{x: win(NP,x\}|:|\{x: Scandinavians(x)\}|) \geq d,$

$\lambda d'.(|\{x: Scandinavians(x)\}$

$\cap \{x: win(gold$

$- medal, x\}|:|\{x: Scandinavians(x)\}|) \geq d, \ldots\}$

c. $\forall Q \in [\![C]\!] [Q \neq \lambda d.(|\{x: Scandinavians(x)\} \cap \{x: win(NP, x\}| :$

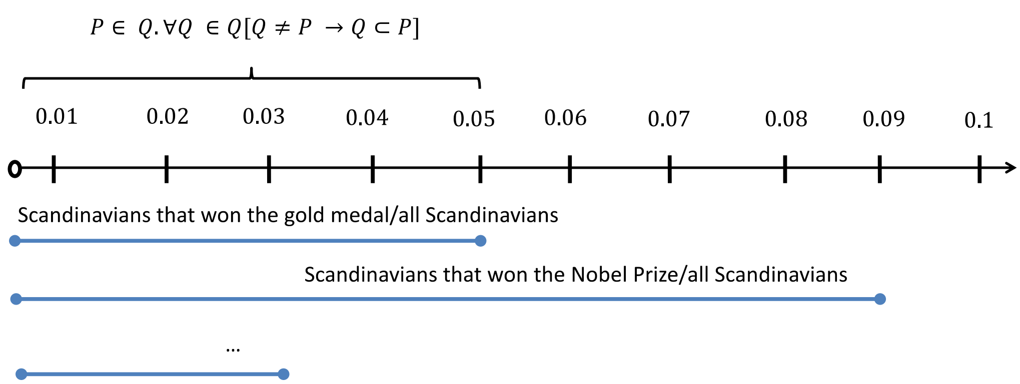
$|\{x: Scandinavians(x)\}|) \geq d \rightarrow$

$Q \subset \lambda d.(|\{x: Scandinavians(x)\} \cap$

$\{x: win(NP, x\}| \; |\{x: Scandinavians(x)\}| :$

$|\{x: Scandinavians(x)\}|) \geq d]$

The utterance and derivation in (44) assumes emphasis on *Nobel Prize*, which according to the account developed here would be licensed by a discourse in which the information structure is structured by a comparison class containing degree sets cataloging all the relevant, resolving alternatives to the Nobel Prize, e.g. other prestigious international prizes like the an Olympic gold medal, and others. This derivation can be visualized as in the number line below.



This account predicts that (44) will be true just in case the proportion of Scandinavians that have won Nobel Prizes is a proper superset of all the other degree sets in the comparison class.[11] Note that this derivation assumes the proportional interpretation of

[11] Given that the proper superset relation entails that the Quantified Set is the uniquely largest set in the Comparison Class, these truth conditions more closely capture the *for the most part* sense of *most* (c.f. Tomaszewicz 2013), rather than the *majority* sense that Romero 2015b uses in her paraphrases. This

many, but that under the regular interpretation, the proportional and the cardinal interpretations are both true in the same situations. Therefore, to fully test the extension of this account to *most*, we must also consider a *most*-utterance with the reverse pattern of Focus/emphasis. However, English does not appear to allow this pattern of Focus-marking; as Romero 2015b states (citing Pancheva &Tomaszewicz 2012 and Tomaszewicz 2013), English does not allow -*est*, in its raised position, to associate with an element inside its base generated DP (see **Figure 2-6**). However, this blocking may be specific to English, because Polish and Bulgarian do allow this association, as the contrast between the acceptability of the paraphrases in (45a) and (46a) illustrates.

(45) John has the best **album by U2** $_F$.

a. # 'John has the better albums by U2 compared to any other band.'

(46) Ivan ima naj-bodri albumin na **U2**$_F$. [Bulgarian]

Ivan has -est-good albums by U2.

b. 'Ivan has the better albums by U2 compared to any other band.'

c. # 'The proportion of albums by U2 that Ivan has is larger than the proportion he has by and other band.'

By extension then, the reverse reading of *most* is in principle possible, just not specifically in English. Moreover, Romero 2015b claims that a sentence like (46) in Bulgarian (and Polish) does not allow for the proportional interpretation paraphrased in (46b), suggesting that *most* is built from a cardinal *many*.

distinction is relevant for interpreting the results of **Experiment 4** in **Chapter 5,** but a full exploration of this distinction is beyond the scope this dissertation.

The account developed in **Section 2.3** above departs from Romero's account in one critical way. By assuming structured meanings as the denotations for licensing QUDs in the discourse, the distinction between the regular, reverse, cardinal and proportional interpretations of *many* are reduced to distinct information structures of distinct discourses. This powerful mechanism therefore allows all these interpretations to arise from one single denotation of *many*, which is assumed to be proportional. Therefore, the account in **Section 2.3** would simply posit an ignorant interlocutor in the discourse in which (46) is uttered. Similarly, this account would imply that discourses that license an utterance like (45) never occur in English, but do in Bulgarian or Polish. Therefore, the account in **Section 2.3** fails to fully capture these observed cross-linguistic differences and the clear role that syntactic mechanisms play in the regular-reverse ambiguity and the cardinal-proportional ambiguity.

A full account of Focus-dependent *most* is beyond the scope of this dissertation. However, the data above demonstrate that *most* provides data that may provide critical constraints on how information structure in the discourse contributes to the derivation of the regular-reverse ambiguity in *many*. Whereas, pragmatic machinery has been the focus of the account developed in this dissertation, *most* demonstrates that pragmatics alone cannot be taken to fully determine the interpretational ambiguities in ambiguous, vague quantifiers. Whatever is responsible for blocking the effects of Focus with the superlative morpheme in English must also be accounted for; and this is unlikely to be a pragmatic factor, as briefly reviewed below.

Following work by Pancheva and Tomaszewicz 2012 and Tomaszewicz 2016, the relative position of the quantifier, degree morpheme, and the Focus operator at LF likely interact to determine the regular-reverse ambiguity. According to Tomaszewicz 2016, superlative adjectives like *most expensive* are ambiguous between the so-called **absolute reading** ((47a)) and two **relative readings** (47b) & (47c) (adapted from Tomaszewicz 2016).

(47) John bought Mary the most expensive cake.

a. John bought Mary **the cake** $_F$ that was more expensive than any other (relevant, alternative) cake. *absolute reading paraphrase*

b. **John** $_F$ bought Mary a more expensive cake than any other (relevant, alternative) person did. *'external' relative reading paraphrase*[12]

c. John bought **Mary** $_F$ a more expensive cake than he did for any other (relevant, alternative) person. *'internal' relative paraphrase*

As the paraphrases in (47) illustrate, the different interpretations arise due to distinct patterns of Focus-marking, just like with *many*, except that the Focus-marking in (47b) and (47c) does not completely rule out the interpretation in (47a).

[12] This term is borrowed from Tieu and Shen 2015, and comes from the fact that Focus is either NP-external or NP-internal, where the NP in question is the object NP, here *the most expensive cake*.

Drawing on cross-linguistic data from Polish, Pancheva and Tomaszewicz 2012 (and Tomaszewicz 2016) claim that definiteness in the modified NP blocks the QR of the degree morpheme -*est* out of its base generated NP. Since Polish does not have definite articles, -*est* can QR outside of the NP, allowing for either NP-internal or NP-external scope, and the ambiguity in (47b) and (47c), unlike English. Thus, in Polish Focus-marking serves to disambiguate between these two possible relative interpretations, and is therefore obligatory in a way that is not observable in English. Therefore, Pancheva and Tomaszewicz 2012 and Tomaszewicz 2016 argue that the scope of -*est* determines which interpretations are possible, and the Focus-marked constituent determines which reading arises (i.e. either John or Mary in (47)). This account likely generalizes to ambiguous, vague quantifiers, helping to explain why the reverse reading may be blocked for *most* in English, but not Polish and Bulgarian. Therefore, a full account of the class of ambiguous, vague quantifiers should consider both the pragmatic factors explored in this chapter and these syntactic factors.

2.4. Implications for accounts of ambiguity and vagueness

Acknowledging that there is still work to be done towards developing a fully unified, comprehensive account of ambiguous, vague quantifiers, the account developed in this chapter does provide some insight into how formal linguistics can address the challenge of adequately capturing both ambiguity and vagueness, by working out a fragment for *many*.

Recall that Kennedy 2019 characterizes ambiguity as the case in which a lexical item can be mapped to multiple, different meanings. Various formal tools are used to capture ambiguity, e.g. scope, type-shifting, or under-specification. Critically, all these approaches attempt to offer principled explanations that predict or constrain the many-to-one mappings between meanings and lexical items; this strengthens the standard semantic assumption that truth-conditions are lexical items can be unambiguous referents to meanings, and that these meanings can be captured by a two-valued logic. In the data reviewed in this chapter, at least four different meanings for *many* were observed. However, in the account developed in **Section 2.3**, these apparently distinct interpretations were instead attributed to distinct interpretations that arise from distinct discourses with distinct information structures. That is, *many* is not actually ambiguous in the sense that this account posits one single semantic denotation, and uses pragmatic machinery to derive these different interpretations. The experiments in **Chapter 3** look for evidence that Focus, via prosodic emphasis, does indeed determine the interpretation of *many* that arises.

Recall that Kennedy 2019 characterizes vagueness as tolerance or boundary-less-ness, giving rise to the Sorites Paradox, in which we find it difficult to identify the particular point at which the second premise becomes false (e.g. if a $5 coffee is expensive, and iteratively a $4.99 coffee is expensive, why do we erroneously accept the conclusion that a $0 coffee is expensive, in the context of this iterative of syllogism). This paradox holds a central role in reasoning about vagueness because it demonstrates this boundary-less-ness, but also because it highlights in turn the potential inadequacy of either a traditional two-valued logic or the assumption of sharp boundaries in capturing vagueness (Kennedy 2011; 2019, citing Fara 2000). For example, if we admit more than two truth-values in our logical system, we can explain how a $3 coffee described as expensive can be super-true but not precisely true, or we could say that a $3 coffee is at the fuzzy, or unknowable edge of the boundary between *expensive* and *not-expensive*. However, supervaluationist accounts still fail to explain how world knowledge can affect judgments of super-truth and truth, i.e. at what point a coffee goes from being expensive to not-expensive. On the other hand, accounts that invoke epistemic uncertainty fail to explain why we continue to accept the second premise of the paradox at each successive iteration, and yet clearly reject the conclusion in isolation.

In the context of ambiguous, vague quantifiers like *many*, according to the account developed in **Section 2.3** there is no change to the standard logical specification of two-valued truth-conditions; *many* itself it not vague, but *many*-utterances appear vague because their interpretation shifts depending on the context and they can be evaluated relative to vague or imprecise contexts. Vagueness is captured by the fact that the comparison class changes depending on the context of utterance. Therefore, there is no one universal boundary, or point where *many* ceases to be true or becomes true – that boundary is defined anew in each context of utterance. Moreover, because the denotation for *many* requires the representations and comparison of cardinalities or proportions, it can be verified via the noisy, approximate representations of magnitude in the ANS (Carey, 2009; Dehaene 1997). The experiments in **Chapter 4** find evidence that this content-dependent aspect of *many* is present and is the default interpretation.

Thus, the semantic-pragmatic account of *many* developed in this chapter is not a complete account of ambiguous, vague quantifiers. However, the innovation of denoting licensing QUDs as structured meanings provides a transparent, fully compositional mechanism that bridges the role of information structure in determining the regular-reverse ambiguity, and for capturing the apparent vagueness of *many*-utterances. The experiments that follow in the next two chapters probe these issues further and suggest further factors that will be integral for a complete account of ambiguous, vague quantifiers.

3. Does Focus determine the interpretation of *many*?: Looking for the regular and reverse interpretation by manipulating Focus and the context

3.1. Introduction

The account laid out in **Section 2.3** leads to the hypothesis that there are two truth-conditionally distinct interpretations *many*-utterances (the regular and reverse interpretations), licensed by the structure and the content of the preceding discourse, which in turn licenses the prosodic Focus-marking; in short, Focus-marking disambiguates between the regular and reverse interpretations of *many*-utterances. For an illustration, consider the data in (1), where emphasis is marked by the subscripted F/CT and bold type-face, the truth-conditions that arise based on this Focus-marking are paraphrased below the sentence (adapted from Herburger 2000, Partee 1988, Romero 2015b, and Westerståhl 1985), and the name for the interpretation is provided in italics.

(1) Many Scandinavians have won the Nobel Prize.

a. Many Scandinavians have won the **Nobel Prize** $_{F/CT}$. *regular proportional*

Paraphrase: Of all Scandinavians, the proportion that have won the Nobel Prize in Literature is larger than the proportion of them that have won other things.

b. Many **Scandinavians** $_{F/CT}$ have won the Nobel Prize. *reverse proportional*

Paraphrase: Of all the people that have won the Nobel Prize, the proportion of them that have been Scandinavians is larger than the proportion of winners from other countries.

According to the account in **Section 2.3**, a particular discourse uniquely gives rise to either (1a) or (1b) via distinct QUDs (the question under discussion, or the point of the discourse; Büring 2003; Ginzburg 1995; Roberts 1996/2012). The information in any discourse is structured by the (successive) QUD(s) in that discourse (Büring 2003; Ginzburg 1995; Roberts 1996/2012) and the goals of the interlocutors (Ginzburg 1995); this information structure licenses, or gives rise to, specific Focus patterns (e.g. Krifka 1995; Rooth 1992; Schwarzschild 1999; von Fintel 1994 for various accounts of this mechanism), which can be prosodically marked by the so-called B-accent (Büring 2003; Jackendoff 1972). In short, the Focus-marking that appears on a particular *many*-utterance signals the specific discourse that would license that particular pattern of Focus-marking and derives distinct truth-values. Therefore, the discourse in which (1a) is uttered is fundamentally distinct from the discourse in which (1b) is uttered; it is in this sense that the two interpretations are thought of as having distinct truth-conditions.

The experiments in this chapter aim to demonstrate the validity of the claim that the information structure of the preceding discourse sufficiently determines the interpretation of a *many*-utterance, as marked by the prosodic B-accent emphasis pattern. To do so, the proportion of the quantified degree set in a contextualizing paragraph was manipulated in such a way that made only either the regular or reverse interpretation of the *many*-utterance true. These experiments measure the rate at which participants select a *many*-

utterance with a Focus-marking pattern that gives rise to truth-conditions that evaluate to true given the facts in the paragraph, and their confidence in that choice. If the account developed in **Section 2.3** is correct in hypothesizing that the distinct truth-conditions of many-utterances arise dependent on the Focus-marking on *many*-utterances and the information structure of the context in which they are uttered, then the results should show that participants systematically select the *many*-utterance with the Focus-marking that evaluates to true, given the facts in the context, and that they are more confident on average when they do so.

In **Experiment 1**, Focus-marking across the two relevant variants of a *many*-utterance was realized by italicizing the word intended to carry prosodic emphasis. Contrary to predictions based on the account in **Section 2.3**, participants were at chance in selecting between the two *many*-utterances. This finding is perhaps not surprising given the numerous different ways that italics can be interpreted (e.g. contrastive emphasis, importance, idiosyncratic usage, technical or definitional usage, etc.), only one of which is the critical B-accent prosody (c.f. Büring 2003) required, by hypothesis, to disambiguate between the regular and reverse interpretations of *many*-utterances. Moreover, Herburger 2000 suggests that the predicate in a *many*-utterance further constrains the interpretive possibilities of *many*-utterances, suggesting that the information structure of the discourse may not be the strongly deterministic cue to interpretation that the account in **Section 2.3** suggests. To address these two possibilities, **Experiment 2** presents a subset of the trials from **Experiment 1**, auditorily, manipulating the prosodic contour on each *many*-utterance, while also manipulating the predicate type in the *many*-utterance. However, the results of **Experiment 2** did not replicate the expected effect of predicate type.

The **Discussion** briefly explores one potential explanation of this failure to replicate, and to find a reliable effect of context, using Optimality Theory (OT; Prince & Smolensky, 1993). Considering a broader range of constraints on the semantic and pragmatic interpretation of *many*-utterances, and how they may interact, a preliminary sketch of relevant constraints, within an OT framework, sheds light on why participants did not show a clear preference for either the regular- or reverse *many*-utterance, depending on the contextual information they received – as the account in **Section 2.3** would predict. This preliminary OT analysis suggests that without any explicit control or manipulation of the interpretation of the subject of the utterance, subjects may have been faced with an interpretive choice, and chose somewhat randomly, thereby decreasing the sensitivity of the experiment to hypothesized effect of context and prosody.

The results of these two experiments and the preliminary OT analysis in the **Discussion** taken together suggest that the semantic, prosodic, and pragmatic factors that determine prosodic Focus-marking, and thereby give rise to either the regular or reverse interpretation of a *many*-utterance, also interact semantic constraints on the interpretation of the subjects in those utterances. In short, the acceptability and felicity of the regular or reverse interpretation depends on more than just the information structure of the preceding discourse; emphasis and information structure cannot fully determine the interpretation of a *many*-utterance. Therefore, in addition to these factors that the account in **Section 2.3** highlights, the results of these experiments motivate further investigations of how the information structure may also determine the interpretive possibilities of plural subjects, and how their interpretation interacts with the critical information structure posited in **Section 2.3**.

3.2. Experiment 1: Aims

The aim of this first experiment was to validate the claim made by the account in **Section 2.3**: that there are two, distinct interpretations available for the relative, proportional quantifier *many* and that these two interpretations are derived via a process in which the contents of the preceding discourse context categorically determines the semantic interpretation of a *many*-utterance. The critical claim being tested in **Experiment 1** in this chapter is that truth-conditionally distinct interpretations arise by virtue of emphasis on one of the two possible constituents that can associate with Focus in the composition of *many*-utterances; i.e. there is a regular interpretation emphasis pattern and a reverse interpretation emphasis pattern.

To test for the availability of these two distinct, context dependent interpretations, paragraphs were constructed that contained facts about the proportion of subsets of items (e.g. fish that live in a lake that either hide in deep or shallow locations). These proportions were such that only one of the two possible truth-conditions of a *many*-utterance would evaluate to true given the facts in the paragraph. Using a between-subjects design, such that each participant only ever saw paragraphs that either made the regular or reverse interpretation of *many*-utterances true, participants were asked to choose which of the two possible emphasis patterns on a *many*-utterance they felt best summarized the information in the paragraph, and provide a confidence rating on a 10-

point scale. If there are indeed two available and distinct truth-conditions, that are in fact distinguished via emphasis marking in a *many*-utterance, participants should be above chance and on average more confident at selecting the emphasis pattern that gives rise to the interpretation that evaluates to true given the information in the paragraph. Such a pattern of results will be taken as support for the proposed information structure-dependent account of the regular-reverse ambiguity in **Section 2.3.**

3.3. Method

3.3.1. Participants

Seventy-one undergraduate students (26 males) at Johns Hopkins University participated in this study in exchange for course credit. To ensure that participants were native, or near-native (White & Genesee, 1996) speakers of English, participants were screened using a survey adapted from Freed, Dewey, Segalowitz, and Halter (2004)[13], after completion of the task. Native and near-native speakers were conservatively[14] defined as people who began acquiring English from at least age 3, and who currently speak English at least 70% of the time with both friends and family over the course of 'a typical week' in the preceding 6-month period. This allows for simultaneous bilinguals who have grown up in multilingual households but excludes people for whom English is not their first and dominant language, for example those who only started learning English in school or later, and those who do not use English, even at school. Based on the

[13] An online version of this survey can be seen here: https://forms.gle/haEFoyy4VPpaW5sk6
[14] Based on the extensive literature on second language acquisition, critical periods, and ultimate attainment (c.f. White, (2003) for an overview), this method followed White & Genesee (1996) and assume that near-native English speakers have attained similar underlying representations as monolingual English speakers.

results of this survey, data from eight participants were excluded from analysis. Data from two further participants were omitted from analysis because of technical issues saving their data files. This resulted in data for analysis from 30 participants (10 males) in the Regular condition, and 30 participants (9 males) in the Reverse Condition. This study was approved by the Homewood Institutional Review Board.

3.3.2. Design and Materials

This experiment followed a between-subjects design with the type of contextualizing paragraph (regular-inducing context, reverse-inducing context) as the between-subjects factor. Two dependent measures were recorded: response type (regular- or reverse-marked *many*-utterance) and confidence rating (on a scale from 1-10, with 1 = not at all and 10 = very confident).

In order to ensure the ecological validity of the results of this experiment, the topics of the paragraphs in this experiment were based on instances of *many* and their surrounding context in COCA, a large (560 million word) and balanced corpus comprised of spoken and written American English from news, drama, fiction, popular magazines, and academic texts (Davies, 2008). Only determiner uses of *many* were selected.[15] The first 20 search results containing a determiner use of *many*, excluding two results containing potentially sensitive or triggering themes (sexual identity and gun violence),

[15] Following Herburger (2000), Solt (2015), and others this dissertation distinguishes semanto-syntactically defined determiners and adjectival uses of *many*. It considers only determiner uses here, by applying tests of ordering (determiners must come before adjectives in English, e.g. the_{det} red_{adj} car) and iterability (adjectives but not determiners can typically be iterated, e.g. the_{det} big_{adj} red_{adj} car, vs. *the_{det} that_{det} big_{adj} red_{adj} car). Under the assumption that distinct semantic types would be required for these distinct syntactic distributions, this account assumes that these two uses of *many* are in fact distinct lexical items. However, a fuller analysis of how and whether the account developed in **Section** 2.3 can account for this adjectival *many* is left for future work.

were minimally altered to create a *many*-utterance of the general form "Many X Y," where X is a constituent phrase (e.g. noun, adjective or prepositional phrase) and Y is predicated of X. One version of this sentence was Focus-marked using italicized text, so as to give rise to the regular interpretation of *many* (e.g. "Many X are *Y*."): the **regular response**; the other was marked so as to give rise to the reverse interpretation of *many* (e.g. "Many *X* are Y."): the **reverse response**. Only one word in the focused phrase was emphasized.[16] Next, the surrounding context of each instance of *many* in COCA was reviewed as inspiration for a contextualizing paragraph.

Each of the 20 paragraphs consisted of three sentences: a preamble, and two paraphrases of the truth-conditions of both the regular and reverse interpretation. In order to limit the influence of variable background knowledge between different participants, about different topics discussed in the paragraphs, the preamble provided context, and gave explicit comparison classes for each of the sets mentioned in the following two sentences. Referring back to the paraphrased truth-conditions in (1) the paraphrase of the regular interpretation of *many* is taken to follow the frame "of all the X, some proportion [that counts as many] are X that Y," and the paraphrase of the reverse interpretation of *many* is taken to be "Of all the Y, some proportion [that counts as many] of those Y are X." Simplified versions of these paraphrases were used as the frames for the two main sentences in the paragraphs. To create the paragraphs for the Regular condition, the proportions in each sentence were manipulated such that the larger proportion appeared

[16] In cases where X and Y contained more than one word, the main content word that contrasted with the comparison class as given in the preamble was emphasized. For example, "*heliotrope* trees" contrasted with "different kinds of native trees…on Hawaiian golf courses."

in the regular paraphrase and the smaller proportion appeared in the reverse paraphrase (this pattern was reversed for the Reverse condition). In order to create variety across trials, the magnitude of the proportions varied (with the larger proportions ranging over: 2/3, 3/4, 4/5, 9/10), such that five paragraphs used each proportion. The preamble always appeared first. Two versions of every paragraph were created, one with the regular paraphrase appearing first, and the other with the reverse paraphrase appearing first. The final sentence always began with "But…" to make the paragraph cohere. An example trial screen is shown in Error! Reference source not found.

Figure 3-1. Example of a trial screen in **Experiment 1**, from the Regular condition. The expected response would be "Many surgeries result from *diabetes*," because the information in the paragraph tells us that diabetes (as opposed to other chronic illnesses) are the leading cause of surgeries.

3.3.3. Procedure

This experiment was presented in JavaScript using the jsPsych (de Leeuw, 2015) API library and plugins, and run in a Chrome v68 browser in full-screen mode.

Participants were randomly assigned to either the Regular or Reverse condition. As described above, participants either saw 20 paragraphs that made the regular interpretation true, or they saw 20 paragraphs that made the reverse interpretation true. For each participant, half of the paragraphs appeared with the regular paraphrase as the second sentence in the paragraph, and the reverse paraphrase as the final sentence; for the other half of participants, the reverse paraphrase preceded the regular paraphrase. The presentation of this sentence order was pseudo-randomized across all the trials in each version of this experiment using a Latin square design, so that there were an even number of regular-first and reverse-first trials. Paragraph presentation order was randomized between participants within each block, and block presentation order was counterbalanced across participants. Each trial consisted of the paragraph, at the top of the screen. Below the paragraph, the two response options were shown with radio buttons to the left of each. The confidence rating scale was shown at the bottom of the screen. Trials were split into four blocks of five, and the order of each block was pseudo-randomized across 16 block-order versions.

Participants were instructed to read the paragraph, then decide which *many*-utterance best summarized the information in the paragraph. Participants were explicitly told that the italicization was meant to indicate which word would be emphasized if the sentence were to be said out loud, and they were reminded to make a choice based on the information as stated in the paragraph, rather than their outside, background knowledge, if they had any. Participants were instructed to then indicate on the scale below the two button choices, how confident they were about their choice, on a rating scale with 10 points displayed as radio buttons (with 1 = “not at all confident” and 10 = “very confident”). Participants had to make a response and give a rating before they could click the ‘continue’ button at the bottom of each screen. In between each block of five trials, participants were shown a simple maze or a visual search task, to encourage them to take a break. Participants were allowed to move at their own pace, taking on average 20 minutes to complete the whole task. During the debrief, participants were informed that the information in these paragraphs had been manipulated, and therefore should not be taken as fact.

3.3.4. Predictions

The account of *many* in **Section 2.3** posits that there are two available interpretations for *many*-utterances and that these two interpretations arise by virtue of the information structure of the preceding discourse context, signaled via information structure-licensed Focus-marking. In short, different truth-conditions arise in *many*-utterances depending on the emphasis (orthographically marked, here) that the utterance (or sentence) carries. If there are in fact different truth-conditions, results should show that participants select the version of the *many*-utterance that evaluates to true given the facts in the contextualizing paragraph they read, i.e. participants in the Regular condition should show systematically higher regular interpretation selection rates, and participants in the Reverse condition should show systematically lower regular interpretation selection rates.

Confidence rating responses are predicted to mirror the selection responses. Thus, higher ratings on average are predicted when participants do select the interpretation that is congruent with their condition.

3.4. Results

Response type – whether the participant selected the regular or reverse interpretation on a trial – was recorded for each trial for each participant. From these raw responses, a regular response variable was then coded (1 = regular response, 0 = reverse response), from which a regular response rate was calculated. Recall, that the rate of selecting the regular response was predicted to be significantly higher in the Regular condition. In the Regular condition, the mean regular response rate was 0.523 ($SD = 0.451$), and in the Reverse condition it was 0.471 ($SD = 0.471$). While the magnitude of the rates is in the expected direction, the fact that they are both so close to chance with such large standard deviations suggests that there is no reliable distinction between the rate at which participants across the two conditions chose the regular response. A logistic regression with condition as the fixed factor and subject and trial as random factors corroborates this finding: there is no significant difference in regular response rates between the two conditions ($p = 0.244$). These mean regular response rates, then, do not support the hypothesis that distinct truth-conditions arise by virtue of emphasizing different elements of a *many*-utterance.

Participants also provided a confidence rating on each trial, indicating how confident they were about the response choice they made. Recall that mean ratings were predicted to be higher on trials where participants selected the response that would evaluate to true given the facts in the paragraph (the congruent response). Therefore, as shown in **Figure 3-3,** mean rating values split by condition were compared to the average rating value on trials where participants selected the regular or reverse response.

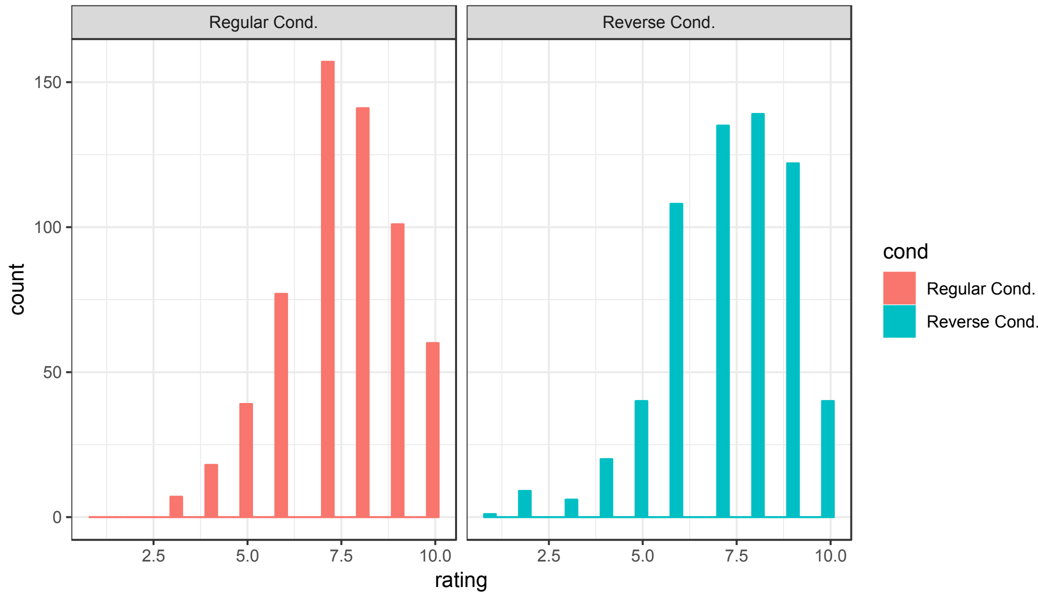


Figure 3-2. A histogram showing the distribution of confidence rating responses made by participants in **Experiment 1,** across both conditions. The mean rating was 7.382 and the median was 7.

As **Figure 3-2** shows, rating values ranged from 1 to 10, and were normally distributed but shifted towards the upper range of the scale (centered at 7 instead of 5). Referring to **Figure 3-3**, these differences in ratings appear to depend on which condition the participant was in, with participants in the Regular condition showing little difference in their confidence ratings regardless of which response they selected (regular response: $M = 7.567$, $SD = 1.532$; reverse response: $M = 7.378$, $SD = 1.584$), but participants in the Reverse condition showing higher confidence ratings when they chose the reverse response ($M = 7.540$, $SD = 1.638$) compared to the regular response ($M = 7.010$, $SD =$ 1.678).). When participants chose the regular response, their mean rating value was numerically smaller ($M = 7.299$, $SD = 1.627$) compared to when they chose the reverse response ($M = 7.464$, $SD = 1.614$). There was also a difference in mean rating value between the conditions, with lower mean ratings in the Reverse condition (*M* = 7.290, *SD* = 1.676) compared to the Regular condition (*M* = 7.477, *SD* = 1.559).

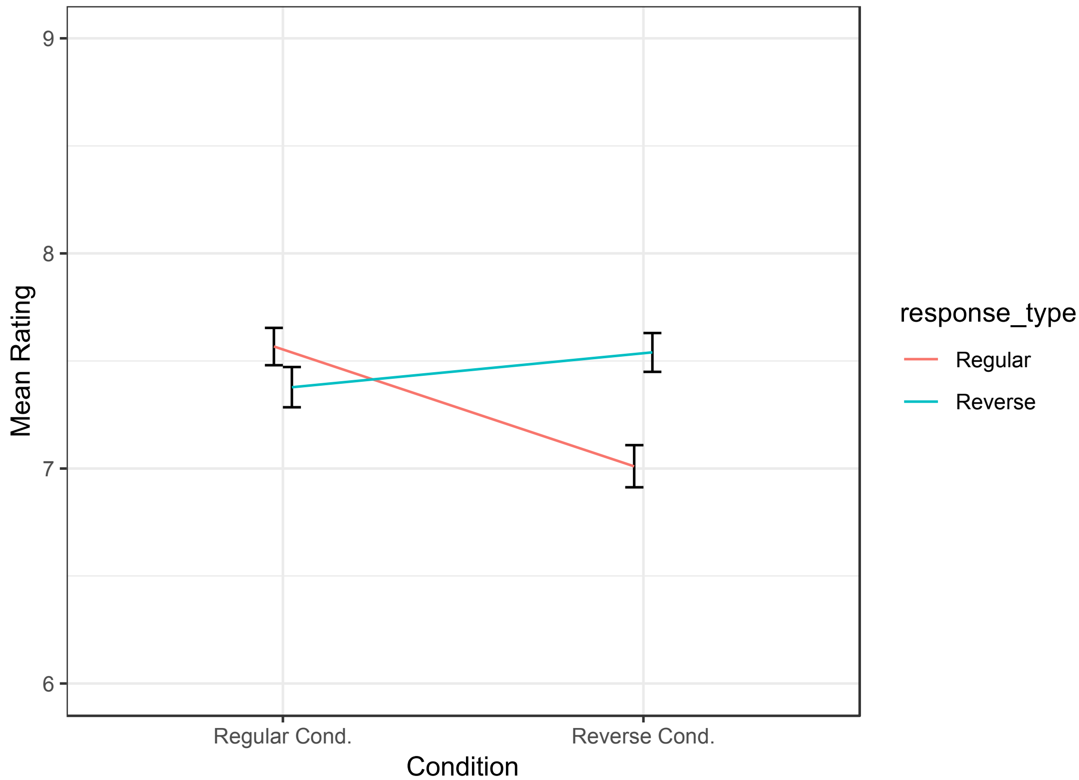


Figure 3-3. Mean rating values split according to whether the participant selected the regular or reverse response on that trial, across the Regular and Reverse conditions in **Experiment 1**. Error bars represent standard error.

To test whether these numerical differences are statistically significant, these data were entered into an ordinal logistic regression with condition and response type, as well as their interaction term, regressed on the rating value, with both participant and item as random effects, using the *ordinal* package in R (v3.5.1). Based on the results of this model, the interaction between condition and response type was significant ($\beta = 0.562$, $z = 2.487$, $p = 0.013$); neither the difference in mean rating values depending on response type ($p = 0.839$), nor condition ($p = 0.180$) were statistically significant. This corroborates what can be observed by inspecting **Figure 3-3**: the difference in rating value between response types depends on the condition, with this effect being driven by a difference in mean rating value in the Reverse condition. So, these results only partially support the hypothesis: participants were on average more confident when they chose the congruent response, but only in the Reverse condition.

While this pattern of response choices and rating values are not wholly consistent with the predictions, it does offer some support for the hypothesis that there are two possible interpretations for *many* that arise in a context-dependent way.

3.5. Experiment 1: Discussion

Contrary to predictions, there was no statistically significant difference in the rate at which participants chose the regular response across the two conditions. However, participants' confidence ratings were significantly lower when they chose the incongruent response, but only in the Reverse condition, partially confirming the predictions. Participants' confidence did seem to distinguish between a congruent and incongruent response, but only when attempting to summarize facts that made the reverse interpretation of *many* true. However, it is possible that participants were not able to reliably distinguish the unique meanings communicated by the regular and reverse response simply because the Focus-marking manipulation – using just italicized text – was not powerful enough. Italicized text is used to convey a number of things in written text, including importance, but also technical, formal, or idiosyncratic meanings, loanwords and much more. Therefore, it is possible that some participants on some (or all) trials were unable to reconstruct the intended meaning of the *many*-utterances, and therefore made selection responses based on unintended meanings.

The fact that participants' confidence ratings only differed in the Reverse condition suggests, however, that when the facts in the context make the reverse response the intended interpretation, participants can use emphasis to distinguish the two response options, but when the facts in the context make the regular response the intended interpretation, emphasis is not a sufficient cue to distinguish these two response options. Thus, emphasis, and perhaps information structure itself (modulo the fact that emphasis can mark other things besides Focus, c.f. Hendriks 2004, a.o.), are not the categorical determinants of the interpretation of *many* that the account in **Section 2.3** predicts.

3.5.1. Additional determinants of the regular-reverse ambiguity

In her account of the interpretive possibilities of *few*, Herburger 2000 observes that the reverse interpretation can only surface with stage-level predicates, and suggests that this may be related to semantic factors that determine the interpretive possibilities of plural subjects with certain predicate types, and syntactic factors that constrain how focus interacts with these interpretations. For example, a specific-generic ambiguity arises with plural subjects (e.g. *Many students publish poetry*.). According to Diesing (1988; see also Carlson 1977) bare plural NPs that are subjects of stage-level predicates (e.g. *stampeded, think, eat*, etc.) can be interpreted either specifically or generically (e.g. (2)). In contrast, bare plural NPs that are subjects of individual-level predicates (e.g. *live, be*, etc.) can only be interpreted generically (e.g. (3)).

(2) Emanuel Feuermann played the cello.

a. **Generic interpretation**: in the general case, Emanuel Feuermann played the cello. (i.e. cello playing defines general case of EM)

b. **Specific interpretation**: there was a particular instance we have in mind in which Emanuel Feuermann played the cello.

(3) Albino rats have red eyes.

a. **Generic interpretation**: in the general case, albino rats have red eyes. (i.e. having red eyes defines the general case of albino rats).

b. **Specific interpretation**: # there is a particular set of albino rats that have red eyes.

Diesing's account, critically relies on whether a predicate's subject is base generated in the 'inner' or 'outer' subject position at LF. Diesing draws a comparison between these two possible subject positions and the tripartite structure of quantifiers (e.g. *Q*(*domain*, *scope*); as well as the ~ operator, e.g. ~(*focus*, *background*), see Rooth 1992), and claims that stage-level predicates allow both the generic and specific interpretation because their subjects are base generated in the 'inner' position (associated with the specific reading) and can move up to the 'outer' position (associated with the generic reading), whereas the subjects of individual-level predicates are base generated in the 'outer' position. Diesing further claims that Focus can only project from the 'inner' position, thereby predicting that only stage-level predicates can have Focus-affected readings (i.e. the reverse interpretation of *many*).

The account in **Section 2.3** does not address the generic-specific ambiguity in plural subjects at all. However, since the sister phrases of *many* are plural NPs in the *many*-utterances in this experiment, this kind of ambiguity could be interacting in the current set of stimuli. If this is the case, that would mean that the reverse interpretation is grammatically illicit for any *many*-utterances containing an individual-level predicate subject (e.g. *Many holes are defined by heliotrope trees*, eight of the 20 trials). To investigate whether this could be an alternative explanation of the results found in **Experiment 1**, each *many*-utterance was categorized as either an individual- or stage-level predicate, by two trained linguists[17], using the following definitions taken from Herburger 1997 (p. 55; adapted from Carlson 1977 and Milsark 1977):

> **Individual-level predicates**: some trait possessed by the entity and which is assumed to be more or less permanent, or at least to be such that some significant change in the character of the entity will result if the description is altered.
>
> **Stage-level predicates**: conditions in which an entity finds itself and which are subject to change without there being any essential alteration of the entity.

Using this new predicate-type variable, the response and rating data from **Experiment 1** were re-analyzed. The regular response rate was entered into a logistic regression with condition and predicate type, as well as their interaction, as fixed effects and participant and trial as random effects. Based on this model, as **Figure 3-4** shows, there was a significant interaction between condition and predicate type (β = 0.894, z =

[17] The two coders agreed on all but one utterance (*Many surgeries result from diabetes*.); this was coded as an individual level-predicate for the purposes of this analysis, reasoning that it is an intrinsic fact about diabetes that it causes many conditions which require surgery.

3.034, p = 0.002) driven by the difference in regular response rates on trials with individual-level predicates which depended on the condition, whereas response rates did not differ depending on condition for trials with stage-level predicates. Moreover, participants in the Reverse condition selected the regular response at a significantly lower rate overall (M = 0.471, SD = 0.500) compared to those in the Regular condition (M = 0.523, SD = 0.500) (β = -0.964, z = -2.776, p = 0.006). In addition, the regular response rate is significantly lower within stage-level predicates (M = 0.461, SD = 0.498) than individual-level predicates (M = 0.579, SD = 0.494; β = -1.234, z = -2.480, p = 0.013).

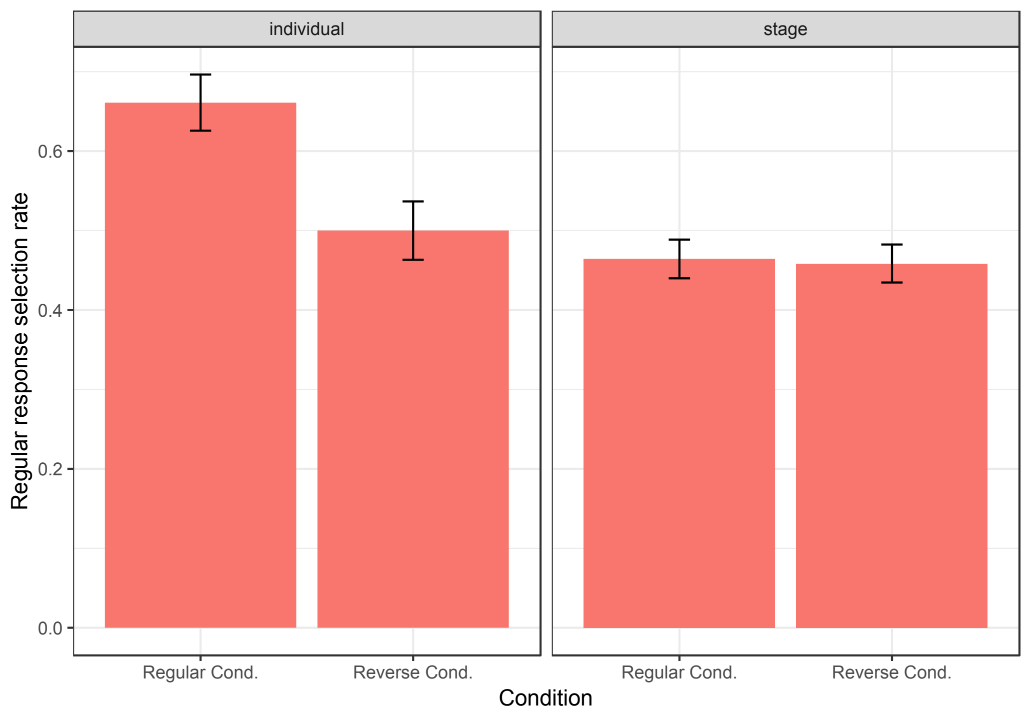


Figure 3-4. Mean regular response rates in **Experiment 1**, split by the type of predicate type in the *many*-utterance, in each condition. Error bars represent standard error.

Confidence rating values were also re-analyzed with the new predicate-type variable using an ordinal logistic regression with condition, response type, and predicate type, as well as all their two- and three-way interaction terms, with both participant and item as random effects. The three-way interaction was not significant, so it was removed and the model was re-run. The results of the model corroborate the pattern that is evident in **Figure 3-5**. As before, the interaction between the condition and the response type was significant ($\beta = 0.565$, $z = 2.471$, $p = 0.013$). This suggests that participants in the Reverse condition were significantly more confident when they chose the reverse response ($M = 7.540$, $SD = 1.638$) compared to when they chose the regular response ($M = 7.010$, $SD = 1.678$), and compared to participants in the Regular condition, who were more confident when they chose the regular response ($M = 7.567$, $SD = 1.533$) compared to when they chose the reverse response ($M = 7.378$, $SD = 1.584$), regardless of the kind of predicate type. Neither the interaction between condition and predicate type ($p = 0.914$), nor between response and predicate type ($p = 0.147$) were significant. In addition, there was no significant difference in mean rating values depending on whether the regular response was chosen ($M = 7.300$, $SD = 1.627$), compared to the reverse response ($M = 7.465$, $SD = 1.614$; $p = 0.364$); between the Regular condition ($M = 7.477$, $SD = 1.559$) and the Reverse condition ($M = 7.290$, $SD = 1.676$; $p = 0.235$); nor between trials with stage-level predicates ($M = 7.389$, $SD = 1.602$) compared to individual-level predicates ($M = 7.366$, $SD = 1.669$; $p = 0.527$).

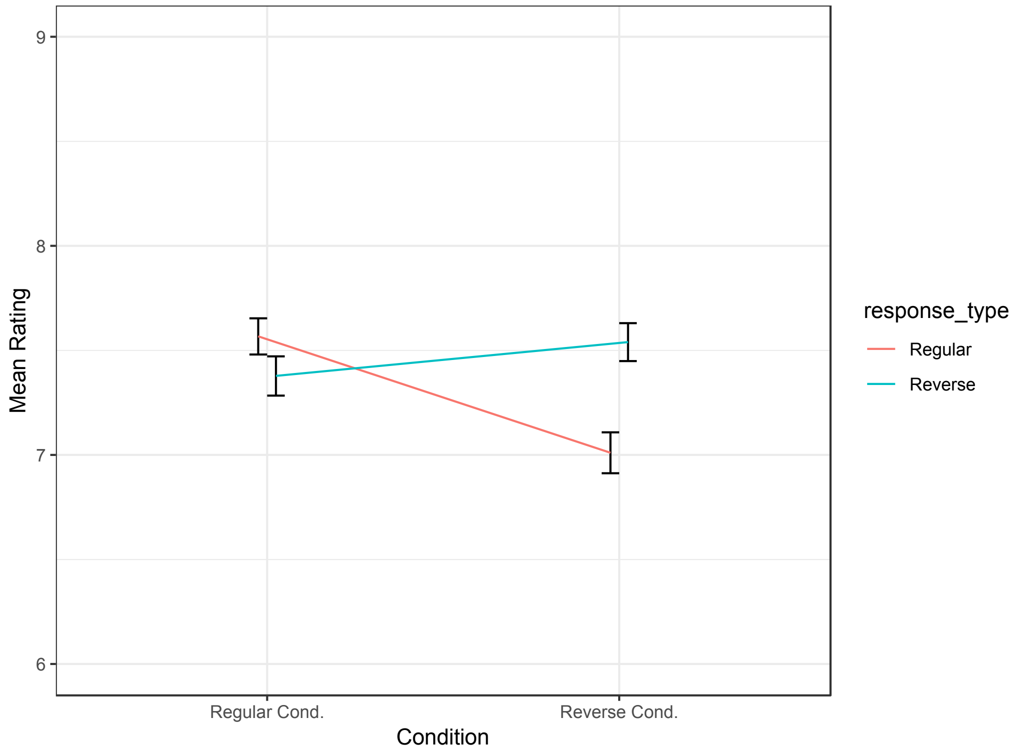


Figure 3-5. Mean rating values from **Experiment 1** split by condition, comparing the average rating value on trials where participants selected the regular or reverse response. Error bars represent standard error.

In short, the semantic constraints introduced by the difference between stage- and individual-level predicates do appear to play some role in participants' ability to select between the *many*-utterances on a trial, but not in how confident they were about that selection. Thus, selection rates may well reflect the combined effect of semantic, prosodic, and pragmatic constraints on the regular-reverse ambiguity in *many* (based on the interaction between condition – the pragmatic factor – and predicate type – the semantic factor). More subtly, confidence ratings reflect the influence of contextual constraints, but only when the response choice being rated adhered to the prosodic and pragmatic factors hypothesized to determine the interpretation of that response.

Based on the possible import of prosody and the predicate-type considerations, **Experiment 2** entertains the hypotheses that: (i) auditory prosodic contours are critical for the accurate reconstruction of the intended interpretation of a *many*-utterance; and (ii) the semantic constraints hypothesized to play a role in plural subjects may also be interacting with the originally hypothesized effect of the information structure of the discourse context (provided in the paragraphs). Each of these hypotheses is addressed in **Experiment 2** in an attempt to tease apart the effects of pragmatically determined Focus, semantic constraints from the predicate type, and prosody on the interpretation of *many*.

3.6. Experiment 2: Aims

Experiment 2 addressed the limitations identified in the materials of **Experiment 1**, and further probed the hypothesis that the Focus-marking involved in the regular-reverse ambiguity is a particular prosodic contour (Büring's 2003 B-accent) that cannot unambiguously be signaled typographically. **Experiment 2** also investigated the hypothesis that the influence of Focus interacts with semantic constraints introduced by predicate type in the *many*-utterance.

Experiment 2 presented both the context and the alternative *many*-utterances auditorily, while balancing the number of stage- and individual-level predicates across trials. With clearer prosodic Focus-marking, intuitive judgements about the meaning of each *many*-utterance seemed more robust, and so a more robust effect of context was predicted (i.e. the condition manipulation). The post-hoc analysis of data from **Experiment 1** and Herburger's 2000 account, predict that with the predicate type carefully manipulated here, an effect of predicate type should interact with the prosody/context manipulation, such that a difference in selection rate should only arise where the semantics of the predicate allows the regular-reverse ambiguity to arise (i.e. stage-level but not individual-level predicates).

Experiment 2 used a subset of trials from **Experiment 1**, balanced for stage-level and individual-level predicates, and presented both the same contextualizing paragraphs and *many*-utterances auditorily. All other aspects of the task remained the same.

3.7. Method

3.7.1. Participants

Sixty-three members of the Johns Hopkins University community with normal or corrected to normal vision participated, and were paid $5 in exchange for participating. The participants in this experiment also participated in **Experiment 4** (see **Chapter 5**), always participating in **Experiment 4** before participating in **Experiment 2**. Thirty-two people (12 males) participated in the Regular condition and 31 people (14 males) participated in the Reverse condition. To ensure that participants were native, or near-native (White & Genesee, 1996) speakers of English, participants were screened using the same survey as in **Experiment 1**, after completion of the task. The participation criteria were the same; based on these criteria, three participants' data were excluded from analysis. This procedure was approved by the Homewood Institutional Review Board.

3.7.2. Materials and Design

Ten trials from **Experiment 1,** indicated in bold in **Appendix A: Experiment 1 & 2 Stimuli,** were re-used in this experiment: five trials had *many*-utterances with individual-level predicates and the other five had stage-level predicates (see the **Discussion** for **Experiment 1** above for how these were categorized). Nothing in the paragraphs or the *many*-utterance options was changed. However, in this experiment, participants were required to listen to the pre-recorded audio of each paragraph and each *many*-utterance being read aloud. The contextualizing paragraph was read in a neutral tone of voice, carefully avoiding any emphasis on any particular elements. Each *many*-utterance was carefully articulated with the B-accent falling on the contrastive element (the word italicized in **Experiment 1**). This audio was recorded in a sound attenuated room.[18]

3.7.3. Procedure

As in **Experiment 1**, participants were randomly assigned to either the Regular or Reverse condition, and the same counter-balancing measures were taken to ensure that the order of the sentences within each paragraph and the order of the *many*-utterance options were evenly balanced across participants. With the smaller number of trials there were only two blocks of five trials in each, but the order of these blocks was also counter-balanced across participants, and the order of trials in each block was randomized across participants. In between the two blocks there was a simple maze to encourage participants to take a quick break.

[18] To hear examples of the audio used in the experiment, follow these links:
Paragraph 1(Reg. cond., order 1):
https://drive.google.com/file/d/1IT1gYiOxaEwZs7kPFiXTCAp_iulZzoSk/view?usp=sharing
Regular *many*-utterance 1:
https://drive.google.com/file/d/1CJmxI8GcRrACUQpYa1GoD33DSLKWTroP/view?usp=sharing
Reverse *many*-utterance 1: https://drive.google.com/file/d/1eI9ij5J9_L0jNlpsuvsxkavfF_BQ-L4q/view?usp=sharing

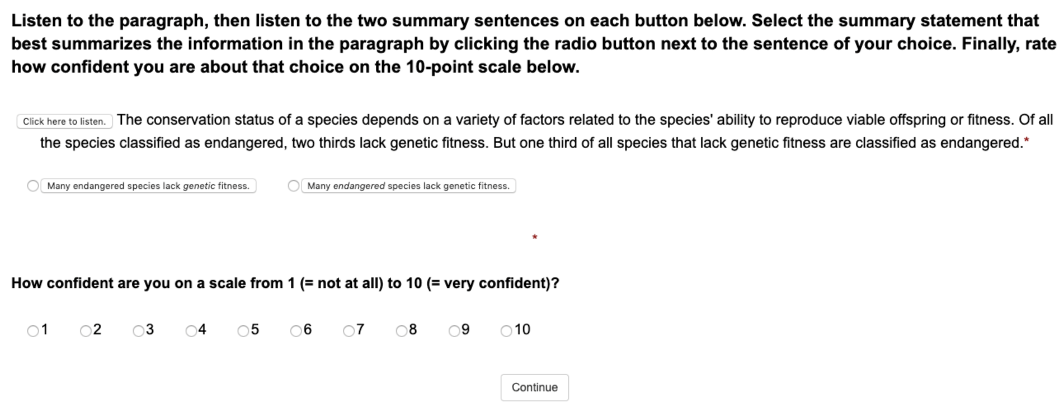


Figure 3-6. An example screen of a trial in **Experiment 2.** Participants could click the html buttons (the grey boxes around the response options and at the start of the paragraph) as many times as they liked to hear the audio for those stimuli. This was the only change to the presentation that was made between **Experiment 1** and **2**.

An example screen from this experiment is shown in **Figure 3-6**. On a trial in this experiment, participants were instructed to first play and listen to the audio recording of the paragraph, and were invited to read along if they wished. After the paragraph audio finished, they were instructed to listen to the audio for each of the two versions of the *many*-utterance in turn. They were allowed to listen to these audio clips as many times as they liked. As in **Experiment 1**, they were then instructed to choose which *many*-utterance they felt best summarized the information in the paragraph, and finally rate how confident they were about that choice on a 10-point scale. This task took participants about 10 minutes to complete on average. During the debrief, participants were informed that the information in these paragraphs had been manipulated, and therefore should not be taken as fact.

3.7.4. Predictions

Based on the new auditory presentation, if prosodic Focus-marking, via the discourse information structure that it signals, serves to sufficiently disambiguate between the regular and reverse interpretation of *many*, then there should be a much clearer, stronger effect of the condition manipulation here compared to **Experiment 1**. That is, participants' selection rates should track closely with the condition they were randomly assigned to.

Moreover, if the availability of the reverse interpretation is blocked by some mechanism that only operates over the semantics of individual-level predicates (Diesing 1988; Herburger 2000; a.o.), but not stage-level predicates, the rate of selecting the regular response should be significantly higher on individual-level trials, especially in the Regular condition, since the context and the grammar both push participants towards this response. In comparison, because stage-level predicates in theory allow both interpretations, the regular response rate should track with the condition the participant is in, but there should be an overall lower regular response rate compared to individual-level predicate trials, given this possible variability. Confidence ratings were expected to again track roughly with congruent responses – when a participant selects the regular response in the Regular condition, they should be more confident than when they select the incongruent response; like response rates, confidence may also see a boost on trials in the Regular condition with individual-level predicates since the regular interpretation is both grammatically and pragmatically licit.

3.8. Results

Response type was coded, and a regular response rate was calculated, as in **Experiment 1**. Referring to **Figure 3-7**, participants in the Regular condition appear to have selected the regular response at a higher rate overall (M = 0.441, SD = 0.498) compared to those in the Reverse condition (M = 0.379, SD = 0.486). Moreover, as predicted, the regular response rate appears to be much higher within individual-level predicates (M = 0.453, SD = 0.499) than stage-level predicates (M = 0.364, SD = 0.482). However, when these data were entered into a logistic regression with condition and predicate type, as well as their interaction, as fixed effects and participant and trial as random effects, the interaction term was not a significant predictor. When the interaction was removed and the model was re-run, neither of these main effects was significant (condition: β = -0.341, z = -0.986, p = 0.324; predicate type: β = -0.439, z = -1.867, p = 0.0619). Thus, unlike data from **Experiment 1**, the type of predicate type does not appear to have any effect on response choice in **Experiment 2**. This unexpected failure to replicate the post-hoc results from **Experiment 1** could either be because: (i) the results of **Experiment 1** are spurious, due to the uneven number of each predicate type in that original stimulus set, or (ii) because the smaller number of stimuli here (only five trials for each predicate type) reduced the statistical power of **Experiment 2**. These two possibilities are discussed further in the **General Discussion** below.

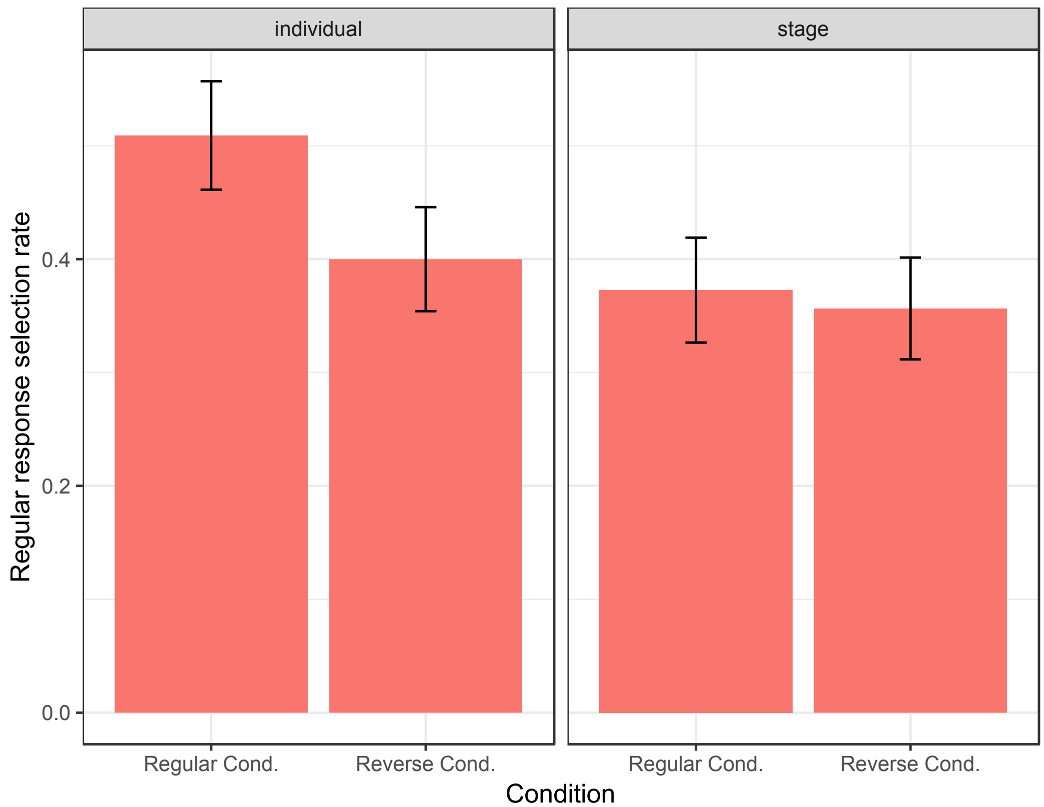


Figure 3-7. Mean regular response rate split by predicate type, across both conditions in **Experiment 2.** Error bars represent standard error.

As in **Experiment 1**, participants again provided a confidence rating on each trial, indicating how confident they were about the response choice they made. Recall that mean ratings were predicted to be higher on trials where participants selected the response that would evaluate to true given the facts in the paragraph, and larger condition-dependent differences were expected on trials with individual-level predicates because both the grammar and the context push them towards the same response. Therefore, **Figure 3-9** considers mean rating values split by condition, comparing the average rating value on trials where participants selected the regular or reverse response, across both predicate types.

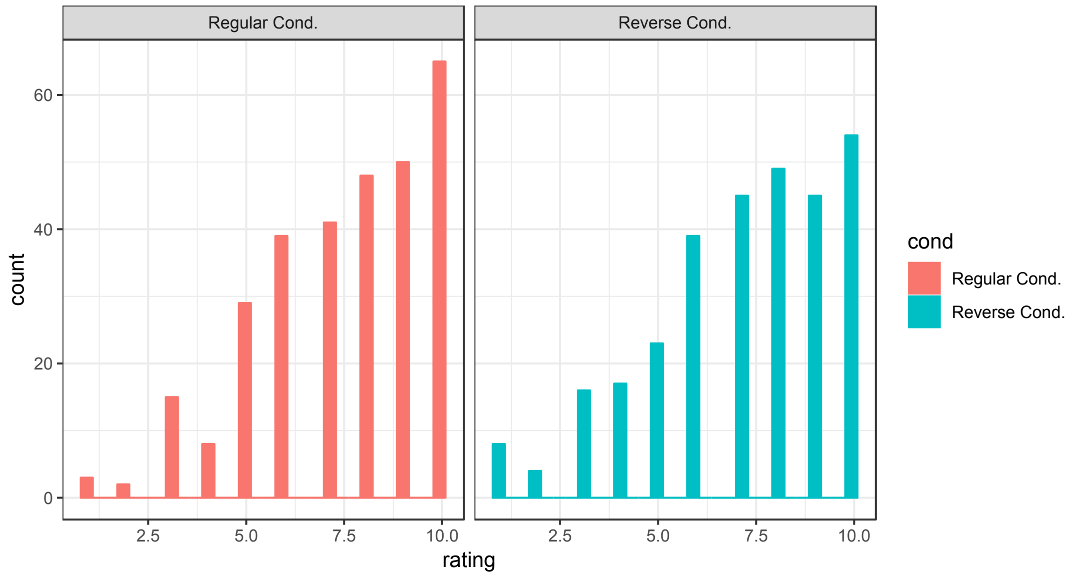


Figure 3-8. A histogram showing the distribution of confidence rating responses made by participants in **Experiment 2**, across both conditions. The mean rating was 7.281 and the median was 8.

As **Figure 3-8** shows, unlike **Experiment 1**, rating values ranged from 1 to 10, but were not normally distributed, skewing toward the upper end of the scale across both conditions (with a mean of 7.281 but a median of 8). Mirroring the results of **Experiment 1**, mean rating values appear lower when the regular response was chosen (*M* = 7.277, *SD* = 2.412), compared to the reverse response (*M* = 7.635, *SD* = 2.020). However, unlike **Experiment 1**, mean rating values were slightly higher in the Regular condition (*M* = 7.595, *SD* = 2.019) compared to the Reverse condition (*M* = 7.386, *SD* = 2.348). Finally, contrary to predictions, mean rating values were also slightly higher on trials with stage-level predicates (*M* = 7.533, *SD* = 2.214) compared to individual-level predicates (*M*= 7.444, *SD* = 2.177). Considering **Figure 3-9**, these differences in mean confidence ratings appear to depend largely on the much lower mean confidence rating for trials with individual-level predicates in the Reverse condition when participants chose the regular response.

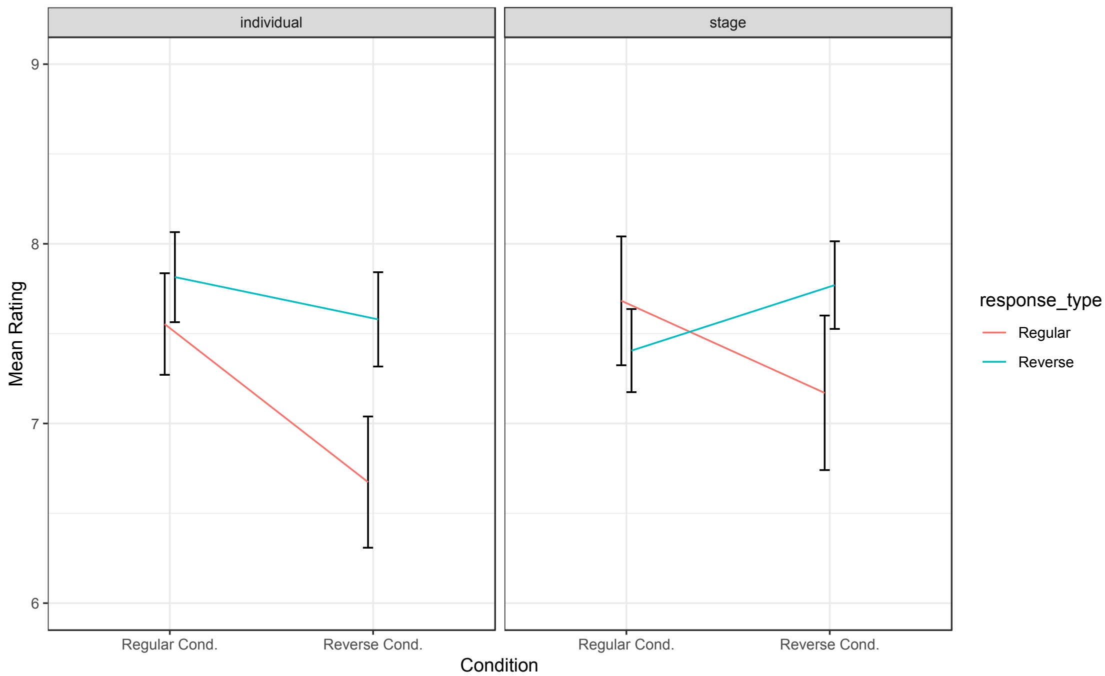


Figure 3-9. Mean rating values split by condition, comparing the average rating value on trials where participants selected the regular or reverse response, across both predicate types for **Experiment 2.** Error bars represent standard error.

To test which of these differences are reliable, the confidence ratings were entered into an ordinal logistic regression with condition, response type, and predicate type, as well as all their two- and three-way interaction terms, with both participant and item as random effects. None of these terms were significant, so the three-way interaction term was removed and the model was re-run. Based on the results of this second model, neither the interaction between predicate type and response type ($\beta = -0.293$, $z = -0.786$, $p = 0.432$), nor condition was significant ($\beta = 0.541$, $z = 1.548$, $p = 0.122$), however, as in **Experiment 1** the interaction between response type and condition was ($\beta = 0.913$, $z =$

2.248, p = 0.025). There was no simple effect of condition (β = -1.101, z = -1.903, p = 0.057), response type (β = 0.106, z = 0.324, p = 0.746), nor predicate type (β = -0.012, z = -0.032, p = 0.975). These results suggest that just as in **Experiment 1**, participants were more confident when they chose the response congruent with their condition, especially participants in the Reverse condition, regardless of the predicate type.

3.9. Experiment 2: Discussion

To summarize, **Experiment 2** used auditorily presented *many*-utterances in order to more directly test the hypothesis that the regular-reverse ambiguity in *many* is disambiguated by context-dependent Focus-marking. **Experiment 2** also manipulated the type of predicate contained in the *many*-utterances to further investigate the role of these semantic factors, and whether they interact with pragmatic factors, in determining this regular-reverse ambiguity. While a more robust effect of context (i.e. condition) was predicted in **Experiment 2** based on this auditory presentation, a reliable difference in either selection rates or confidence ratings was not found as a function solely of condition, suggesting that prosody, and therefore possibly the preceding discourse context is not the strongly deterministic disambiguating influence that the account in **Section 2.3** predicts. Alternatively, prosody may not be the best or clearest way to indicate information structure-derived Focus (an abstract marker of new information). Indeed, Hendriks (2004) notes that emphasis is not always a reliable indicator of Focus.

Accepting that Focus may not have a one-to-one mapping with the particular prosody used in **Experiment 2**, this fact alone cannot explain the results in **Experiment 2**, because paragraphs containing contextualizing information and the predicate type in the *many*-utterance choices were both manipulated. These paragraphs provided the information structure of a proceeding discourse. Therefore, participants should have had enough information to reconstruct the intended information structure that would license the intended Focus-marking. Thus, participants received two cues to the intended information structure and meaning: context and prosody. Therefore, while some participants, some of the time may have had lapses in attention, and were therefore unable to re-construct the intended meaning, it is hypothesized that both the information structure provided by the context and the prosody are only part of the set of factors that determine or give rise to a particular interpretation of a *many*-utterance in a given context.

The replicated effect of predicate type and condition in the mean confidence ratings across **Experiment 1** and **2**, suggests that semantic constraints, such as how the sister NP of *many* is interpreted (Diesing 1988; Herburger 2000) may inform the degree of acceptability or felicity of a particular interpretation of a *many*-utterance. However, the inconsistent effect of predicate type, and the lack of the effect of condition, in the binary response choice data, suggest that these pragmatic and semantic constraints may not be the only factors that are involved in determining the regular-reverse ambiguity of *many*-utterances. In the **General Discussion** below, this possibility is explored in a preliminary OT-based analysis (Prince & Smolensky, 1993). A fully articulated OT account of the regular-reverse ambiguity in *many*-utterances is beyond the scope of this dissertation, but one account of a closely related phenomenon is reviewed below, which provides the beginnings of an explanation for the experimental results found here.

3.10. General Discussion

Synthesizing across the results from **Experiments 1** and **2**, consistent evidence was found in the mean confidence ratings that the information structure (as provided by the contextualizing paragraph) did have an effect on which pattern of Focus-marking participants used to summarize that information (i.e. whether they selected the regular or reverse utterance). This data alone would suggest that information structure, along with some aspect – semantic or pragmatic – of the regular or reverse interpretation, are strong determinants of Focus-marking and thereby the interpretation of the *many*-utterance that surfaces. However, the conclusion that the information structure is the sole determinant of emphatic marking (and thereby the interpretation of *many*-utterances) is qualified by the absence of the hypothesized context effect in the response rate data from both **Experiments 1** and **2**. The fact that response rates did not appear to differ at all between conditions, in contrast with confidence ratings, could be due to the fact that this effect is not strong enough to be detected by a binary forced choice task, or the effect of the context on Focus-marking is determined by a more complex set of interacting factors, and therefore more easily detected with the gradable confidence rating responses, given the current experimental design.

In addition, an effect of predicate type was found in the response rate data from **Experiment 1** but not **Experiment 2.** The lack of a reliable effect between stage- and individual-level predicates raises the possibility that certain methodological limitations could be compromising the reliability of the results; perhaps there are further factors that need to be controlled or manipulated in order to detect the hypothesized effect of context.

The discussion that follows describes in turn the two possible explanations for the different patterns of context and predicate type effects across response rate and confidence rating data between **Experiment 1** and **2**; it speculates on possible constraints and rankings that could offer useful insight into the true nature of this regular-reverse ambiguity; and, finally, considers how such constraints and their rankings might inform follow-up experiments that could offer stronger evidence for the interactive role of pragmatic, semantic, and prosodic constraints in the interpretation and use of vague, ambiguous quantifiers.

The first possible explanation for why an effect of predicate type was found in **Experiment 1** response rates and not in **Experiment 2**, could simply be that the effect observed in **Experiment 1** was spurious, i.e. this could be an erroneous rejection of the null hypothesis, perhaps because the number of individual- and stage-level predicates was not balanced in **Experiment 1** (only eight of the 20 trials contained individual-level predicates), and once there was an equal number of both predicate types in **Experiment 2**, this effect correctly disappeared. To confirm this explanation, **Experiment 1** should be replicated, while maintaining its larger number of trials (e.g. 10), but also carefully balancing predicate type, prosody and information structure. If the effects remain reliable, this should be taken as further support for the hypothesis that the regular-reverse ambiguity does indeed exist and depends on prosody, information structure, and the lexical semantics of the predicate type.

However, confidence ratings were reliably and consistently lower when participants in the Reverse condition chose the regular utterance across both experiments. In addition, an effect of predicate type was found in **Experiment 1** regular response rates but not in **Experiment 2.** This observation leads to a second explanation of the results: the design of these experiments, in particular **Experiment 2,** is not sensitive enough to detect the effect of predicate type and context, i.e. this could be an erroneous acceptance of the null hypothesis, since there are only five trials with each predicate type in **Experiment 2**. Indeed, the context had a reliable effect on confidence ratings, suggesting that it does play some role in this ambiguity. Thus, if we continue to entertain the hypotheses that both the discourse context and the semantics of the utterance have real effects on how emphasis in a *many*-utterance is realized and therefore the interpretation that surfaces, then it follows that the current experimental design simply failed to detect these effects. This explanation suggests that the manipulation of context and predicate type are simply not the correct parameterization of the factors that indeed underlie this ambiguity. This possibility is tentatively formalized with an OT analysis below.

The goal in the remainder of this discussion is to explore a potential, preliminary[19] OT-based explanation of why the results failed to show an effect in response rates, in particular why they only showed a regular response rate significantly different from chance in the Regular condition on trials with an individual-level predicate.

In the optimizations below, I follow Hendriks (2004) in assuming that the participant has already constructed an information structure from the paragraph and recognized the

[19] This analysis is preliminary because the full set of logically possible output candidates has not been exhaustively listed and harmonically bounded. Therefore, the tableaux below are meant for illustrative purposes only as a demonstration that the constraints developed based on the relevant literature could in principle yield an explanation of the results of **Experiments 1** and **2**.

utterance, i.e. read the sentence with italics or heard speech sounds, and assigned a global syntactic structure to the *many*-utterance response choice being considered; together these form the input. Thus, the discussion and tableaux below are presented as uni-directional optimizations where the input is taken to be the information provided on a trial: the licensing QUD (reflecting the content of the contextualizing paragraph), the surface syntax of the *many*-utterance, along with the lexical semantics of the quantifier, the Focus operator and the predicate type (individual- or stage-level), as well as the prosodic contour of the *many*-utterance presented (marked with boldface in the tableaux below). Correspondingly, the output is taken to be pragmatic-semantic interpretations: the meaning that participants thought was the best way to summarize the information in their paragraphs. Thus, each tableau below explores the possible optimization process a participant undertook as they decided to select either the regular or reverse utterance, with either a stage- or individual-level predicate, given the information in the contextualizing paragraph.

Hendriks (2004) has developed an OT-based analysis of the focus-sensitive interpretation of sentences containing the quantificational determiner *only* and contrastive emphasis (e.g. (4) below). *Only* is both a determiner with quantificational structure (i.e. two arguments forming the domain and scope, respectively), and a focus particle (i.e. with two arguments forming the focus and the background, respectively), and therefore closely related to the phenomenon of interest here (see Herburger 2000). According to the literature on the interpretation of quantificational determiners like *only*, the N' in their sister phrase (e.g. *ships,* in (4) below) restricts the domain of quantification (i.e. saturates their first argument) while the rest of the clause forms the scope of quantification (i.e. saturates their second argument). According to the literature on the interpretation of the quantificational structure of focus particles like *only* (and, by assumption, operators like ~), only elements in the XP they are adjoined to can form their focal content, all the other elements they c-command must form the background (see also Herburger 2000). This predicts that any focused elements will be in the scope of the quantifier, while any backgrounded information will be in the domain of quantification. Indeed, there is also a noted preference for placing new or contrastive information in the scope of a quantifier. However, there is also a common pattern of emphasizing – referred to more formally as accenting – new or contrasting information (c.f. Büring 2003), resulting in the anaphoric deaccenting of the other material, which should then, by default, go into the background of Focus and the domain of quantification. These semantic and prosodic characterizations of the interpretation of *only* are at odds in sentences like (4). For example, based on the accent pattern in (4) (boldface), *unload* should be in the domain of quantification, but according to the syntax *ships* should be.

(4) Only ships unload **at night**.

Hendriks (2004) proposed four constraints to account for the interpretive possibilities of sentences containing *only* and prosodic emphasis. This analysis adopts three of Hendriks' (2004) constraints, and propose two new constraints largely motivated by the data presented **Chapters 1** and **2**. The first new constraint posited here is a Faithfulness constraint that captures the claim from **Chapter 2** that information structure strongly influences the interpretive possibilities of *many*-utterances. The second is a faithfulness constrain that captures the close relationship between prosody and information structure, inspired by Büring 2003. Together these constraints, within an OT framework begin to explain why participants in the Reverse conditions of both experiments were particularly uncertain about selecting the regular *many*-utterance but only showed an elevated regular response rate on trials with individual-level predicates in the Regular condition in **Experiment 1** but not **2**.

In her analysis of *only*, Hendriks proposed DEACCENTING and FOCUSING.

> DEACCENTING (DEACC)
>
> If a constituent is anaphorically deaccented, it must contribute to the domain of quantification of a quantifier.

This constraint captures the fact that if an element carries contrastive accent, its sister is necessarily anaphorically deaccented (distinct from default accent or no accent; see Büring 2003; Hendriks 2004; reminiscent of the interaction between Schwarzschild's 1999 GIVENNESS and AVOIDF constraints). Accenting and deaccenting go hand in hand, one cannot occur without the other. Therefore, according to the formulation of this constraint, a violation is incurred each time the sister of the accented element does not contribute to the domain of the quantifier. For example, the first output candidate in **Tableau 1** incurs a violation because *win*, the sister of the accented phrase *Nobel Prize*, should be in the domain of *many*.

FOCUSING[20]

> If a constituent contributes to the focus of a focus particle, use this constituent to restrict the scope of the quantification, and use the rest of the clause to restrict the domain of quantification.

This constraint captures the communicative preference against interpreting new or salient, focused, information as contributing to the domain of quantification, since the domain can be thought of as similar to the background or given information, from which some subset is being described by the quantifier (e.g. Scandinavians out of all people who have won Nobel Prizes). This constraint is an output-output constraint, because it evaluates one aspect of an output candidate (the interpretation of ~) against another (the interpretation of *many*). A violation is incurred each time an element in the first argument of the focus operator ~ is not also interpreted in the domain of *many*. For example, the first candidate in **Tableau 1** incurs two violations because *Scandinavians* and *win* appear in the focus of ~, but only *win*, not *Scandinavians*, appears in the scope of the quantifier, and *Nobel Prize*, because it is not part of the focus of ~ should appear in the domain of *many* but appears instead in the scope.

To these constraints, a newly formulated constraint motivated by the account in **Chapter 2** is added which describes the (apparently) deterministic role of information structure in this regular-reverse ambiguity:

[20] In Hendrik's 2004, FOCUSING is conceived as markedness constraint on how the interpretation of the focus structure of *only* maps onto the interpretation of its quantificational interpretation. The formulation of this constraint was altered to more generally apply to any focus particle, not only one that is also quantificational.

INFORMATIONSTRUCTURE (INFOSTRUC)

The interpretation of a focus operator must be faithful to the denotation of the licensing QUD in the discourse context.

This Faithfulness constrain captures the notion that one must faithfully interpret the licensing information structure of a given utterance. Formally, one must interpret the value derived by applying the background function (e.g. $\lambda d'.\ (|\{x{:}\,Scandinavians(x)\} \cap \{x{:}\,win(NP,x\}|{:}\,|\{x{:}\,Scandinavians(x)\}|) \geq d$ for the reverse utterance) to the restrictor set as the focus of ~ (i.e. first argument), and the background function of the licensing QUD (e.g., $\lambda d\,.\,(\{|x{:}\,win\ (NP',x)|) \geq d\}$ for the reverse utterance) as the background (i.e. second argument). A violation is incurred for every element in the interpretation of ~ that does not correspond to the denotation of the QUD as described by its structured meaning. For example, given the input in **Tableau 1**, any output candidate with an interpretation of ~ as *~(win & Nobel Prize, Scandinavians)* would incur three violations. Because this analysis assumes this is the highest ranked constraint, any possible candidates that violate this constrain are ruled out.

The next new constraint is motivated by the observation that the lexical semantics of the predicate, and the syntactic movement that results from those semantics, in a *many*-utterance also appears to play a role in determining their interpretive possibilities. In Herburger's 2000 account of the regular-reverse ambiguity, *many* either undergoes QR, which raises the whole DP containing *many* and its sister NP, or *many* undergoes Q-raising, which raises only the quantifier leaving the sister phrase *in situ*, meaning that intermediate elements (like ~) can operate over that sister. Therefore, Herburger predicts that the reverse interpretation of *many* can only surface when *many* undergoes Q-raising (a kind of A-movement introduced by Herburger that allows only the determiner to raise without its sister phrase), but not QR, implying that the reverse interpretation can only surface when QR is blocked, which Herburger suggests is the case in so-called Definiteness Effect environments, of which stage-level predicates are one. Correlatedly, Herburger proposes that anti-Definiteness Effect environments, of which individual-level predicates are one, require scope taking by QR (whereas optional and obligatory Definiteness Effect environments allow either QR or Q-raising). The constraint below formalizes this observation as a constraint on the semantic interpretation of the quantifier, thereby remaining theoretically neutral on the syntactic effects of the predicate type.

> ANTIDEF
>
> Determiners in the subjects of individual-level predicates must undergo QR, and therefore must be interpreted as the background of the quantifier.

This is an OT formulation of Herburger's 2000 observation that the subjects of individual-level predicates, i.e. the sister NP of *many,* must be interpreted as the background, and not the scope of the quantifier. Therefore, a violation is incurred whenever any element of the sister NP appears in the scope, and not the background of *many* in the tableaux below. For example, the tied winners in **Tableau 4** both incur three violations of this constraint because *Scandinavians* is not interpreted in the background, but *be* and *Nobel Prize* are. Note however that this constraint is vacuously satisfied, and therefore omitted, when a stage-level predicate appears in the input.

Finally, both typographically and auditorily emphasized *many*-utterances were presented, following the claim in Büring 2003 that the contrastive accent, or B-accent, signals a particular strategy in discourse, i.e. it signals critical information about what the speaker takes the information structure of the discourse to be. In the specific case of the B-accent it signals that the accented element is new or newer, i.e. contrastive, information. This idea is formalized in the following novel OT constraint:

FAITHPROSODY

If an element carries a B-accent, interpret as contrastive information.

This constraint could be thought of as the interpretive counterpart to constraints that force new or focused information to be marked in pronounced output (in an optimization where interpretations form the input, see the NEW constraint in Choi 1996 or the MARKFOCUS constraint in Lutken 2019; see also GIVEN from Schwarzschild 1999). This constraint is violated when an interpretive output candidate interprets an element other than the one emphasized in the input as contrastive information. For example, this is the critical constraint that eliminates the two next best output candidates in **Tableau 1** (the candidates right above the two winners).

Tableau 1 and **Tableau 2** below show the optimization process that derives the reverse and the regular interpretation of a *many*-utterance with a stage-level predicate in the input, and **Tableaux 3** and **4** show the optimization with an individual-level predicate in the input. Using these six constraints, we can deduce one possible relative ranking from which we can derive a pattern of optimal candidates that helps clarify the response rate results found in **Experiments 1** and **2**.

INFOSTRUC >> FOCUS >> ANTIDEF >> FAITHPROS >> DEACC

Note that this ranking assumes that pragmatic constraints outrank semantic and prosodic constraints. This assumption is consistent with the view developed in **Chapter 2:** that the interpretive possibilities of *many* depend on the information structure of the preceding discourse (but see Herburger 2000 for an exclusively semantic-syntactic account). [21]

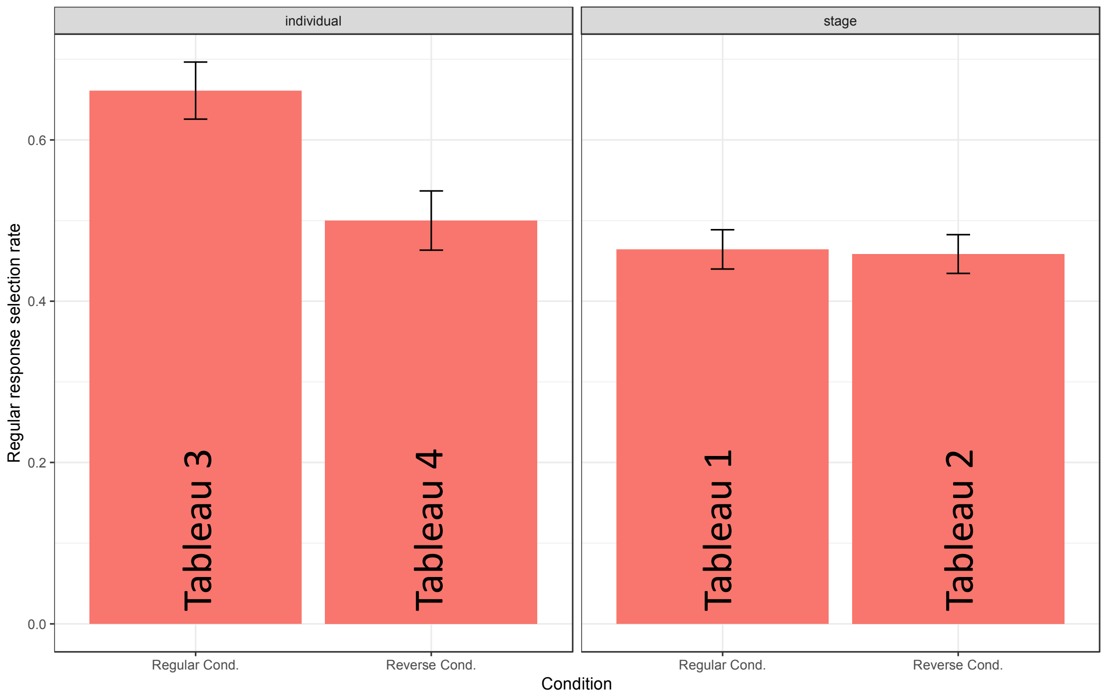


Figure 3-10. Response rate data repeated from **Experiment 1.** Each group of data is labeled according to which tableau in the preceding discussion captures the interpretive possibilities (or lack thereof) that arise for either the regular or reverse *many*-utterance (**Tableaux 1** & **3** or **Tableaux 2** & **4,** respectively), for stage- and individual-level predicate trials (**Tableaux 1 & 2** or **Tableaux 3** & **4,** respectively). In short, the mean rate at which participants' regular selection responses significantly differed from chance is also the only case where the optimization predicts one optimal candidate instead of two (**Tableau 4** compared to **Tableaux 1, 2, & 3**).

Recall that the interaction between predicate type and condition was significant for selection rates in **Experiment 1**, shown in **Figure 3-10.** Assuming that the results of

[21] Indeed, the syntactic constraint is in fact superfluous in these optimizations and has been left in merely to illustrate this point.

Experiment 1 reflect a real effect, and that the null result in **Experiment 2** was due to the smaller number of trials (i.e. a reduction in power), the goal the following optimizations and discussion is to explore why an effect of context was not detected, and what factors can explain the pattern of confidence ratings across both experiments. Each optimization is discussed in more detail below.

Considering the case of stage-level trials first, the tied output winners in **Tableaux 1** and **2** demonstrate that regardless of whether the regular or reverse utterance is selected, the participant is faced with a choice between interpreting the subject of the predicate generically or specifically. Thus, these optimizations suggest that participants may have been at chance in selecting the regular response on these trials because the stimuli in these experiments do not systematically constrain the interpretive possibilities of the subjects; each subject may have made an interpretive choice based on factors that were not controlled or manipulated here. However, these optimizations do not provide a complete explanation of why only participants in the Regular condition showed no difference in their confidence ratings based on which response they chose. If anything, these optimizations suggest that participants in both conditions should have showed flatter or lower overall confidence on stage-level predicate trials.

Tableau 1. This tableau shows how the interpretation of the regular utterance response would be derived. It demonstrates that with a stage-level predicates the interpretive constraints on the quantifier compete with the interpretive constraints introduced by contrastive emphasis, how the emphasis is pronounced and the semantics of the predicate, but that the interpretive constraints of the Focus operator ~ and information structure are the strongest determinants of the interpretation that arises. Recall that the denotation of the licensing QUD posited to give rise to this utterance is $\lambda\ d\ .(\{|x: won(x, Scandinavians)|) \geq d\}$.

Input: QUD: How many Scandinavians have won different honors? Utterance: [[POS C]] [1[t_1-many Scandinavians] have won **NP** $_{F/CT}$]] ~C] (regular *many*-utterance)	INFO STRUC	FOCUS	FAITH PROS	DEACC
many(Scandinavians, win & NP) *~(Scandinavians & win, NP)* Contrasted element: *Scandinavians* Subject: Specific	***!	**	*	*
many(Scandinavians, win & NP) *~(NP, Scandinavians & win)* Contrastive element: *Scandinavians* Subject: Specific		*	*!	*
□ *many(Scandinavians, win & NP)* *~(NP, Scandinavians & win)* Contrastive element: *Nobel Prize* Subject: Specific (regular interpretation)		*		*
many(win & NP, Scandinavians) *~(Scandinavians & win, NP)* Contrastive element: *Scandinavians* Subject: Specific (reverse interpretation)	***!	*	*	
many(Scandinavians, win & NP) *~(NP, Scandinavians & win)* Contrastive element: *Scandinavians* Subject: Generic		*	*!	*
□ *many(Scandinavians, win & NP)* *~(NP, Scandinavians & win)* Contrastive element: *Nobel Prize* Subject: Generic (regular interpretation)		*		*
many(win & NP, Scandinavians) *~(Scandinavians & win, NP)* Contrastive element: *Scandinavians* Subject: Generic (reverse interpretation)	***!	*	*	
etc.				

Tableau 2. This tableau shows how the interpretation of the regular utterance response would be derived. It demonstrates that with a stage-level predicates the interpretive constraints on the quantifier compete with the interpretive constraints introduced by contrastive emphasis, how the emphasis is pronounced and the semantics of the predicate, but that the interpretive constraints of the Focus operator ~ and information structure are the strongest determinants of the interpretation that arises. Recall that the denotation of the licensing QUD posited to give rise to this utterance is $\lambda d\,.(\{|x: win\ (NP', x)|) \geq d\}$.

Input: QUD: How many people of each region have won the Nobel Prize? Utterance: [[POS C]] [1[t_1-many **Scandinavians** $_{F/CT}$] have won NP]] ~C] (reverse *many*-utterance)	INFO STRUC	FOCUS	FAITH PROS	DEACC
many(Scandinavians, win & NP) *~(Scandinavians, win & NP)* Contrastive element: *Scandinavians* Subject: Specific		***!		*
many(Scandinavians, win & NP) *~(win & NP, Scandinavians)* Contrastive element: *Nobel Prize* Subject: Specific (regular interpretation)	*** !		*	*
□ *many(win & NP, Scandinavians)* *~(Scandinavians, win & NP)* Contrastive element: *Scandinavians* Subject: Specific (reverse interpretation)				*
many(Scandinavians, win & NP) *~(Scandinavians, win & NP)* Contrastive element: *Scandinavians* Subject: Generic		***!		*
many(Scandinavians, win & NP) *~(win & NP, Scandinavians)* Contrastive element: *Scandinavians* Subject: Generic	***!			*
many(Scandinavians, win & NP) *~(win & NP, Scandinavians)* Contrastive element: *Nobel Prize* Subject: Generic (regular interpretation)	***!		*	*
□ *many(win & NP, Scandinavians)* *~(Scandinavians, win & NP)* Contrastive element: *Scandinavians* Subject: Generic (reverse interpretation)				*
etc.				

Turning to individual-level predicates, **Tableaux 3** and **4** demonstrate the importance of the anti-Definiteness Effect environments, since the ANTIDEF constraint eliminates the interpretive ambiguity of the plural subjects on trials with individual-level predicates, but critically only when the input contains the regular QUD. Thus, these optimizations help explain why only participants in Regular condition showed significantly higher selection rates on trials with *many*-utterances that contained an individual level predicate. Specifically, this analysis and participants' responses in these tasks suggest that, contra Herburger 2000, individual-level predicates are not absolute anti-Definiteness Effect environments; ANTIDEF only blocks the generic-specific ambiguity of plural subjects of individual-level predicates when the information structure of the discourse gives rise to a reverse interpretation.

Tableau 3. This tableau shows how the interpretation of the regular utterance response would be derived. It demonstrates that with an individual-level predicates the interpretive constraints on the quantifier compete with the interpretive constraints introduced by contrastive emphasis, how the emphasis is pronounced and the semantics of the predicate, but that the interpretive constraints of the Focus operator ~ and information structure are the strongest determinants of the interpretation that arises. Recall that the denotation of the licensing QUD posited to give rise to this utterance is $\lambda\, d\, .\, (\{|x: be(x, Scandinavians)|]) \geq d\}$.

Input: QUD: How many Scandinavians have each different hair color? Utterance: [[POS C]] [1[t_1-many Scandinavians] are **blonde** $_{F/CT}$]] ~C] (regular *many*-utterance)	INFO STRUC	FOCUS	ANTI DEF	FAITH PROS	DEACC
many(Scandinavians, be & blonde) *~(Scandinavians & be, blonde)* Contrastive element: *Scandinavians* Subject: Specific	***!	**		*	*
many(Scandinavians, be & blonde) *~(blonde, Scandinavians & be)* Contrastive element: *blonde* Subject: Specific (regular interpretation)		** !			*
many(be & blonde, Scandinavians) *~(Scandinavians & be, blonde)* Contrastive element: *Scandinavians* Subject: Specific (reverse interpretation)	***!	*	***	*	
many(Scandinavians, be & blonde) *~(Scandinavians & be, blonde)* Contrastive element: *Scandinavians* Subject: Generic	***!	**		*	*
many(Scandinavians, be & blonde) *~(blonde, Scandinavians & be)* Contrastive element: *Scandinavians* Subject: Generic		**		*!	*
□ *many(Scandinavians, be & blonde)* *~(blonde, Scandinavians & be)* Contrastive element: *blonde* Subject: Generic (regular interpretation)		*			*
many(be & blonde, Scandinavians) *~(Scandinavians & be, blonde)* Contrastive element: *Scandinavians* Subject: Generic (reverse interpretation)	***!	*	***	*	
etc.					

Tableau 4. This tableau shows how the interpretation of the regular utterance response would be derived. It demonstrates that with an individual-level predicates the interpretive constraints on the quantifier compete with the interpretive constraints introduced by contrastive emphasis, how the emphasis is pronounced and the semantics of the predicate, but that the interpretive constraints of the Focus operator ~ and information structure are the strongest determinants of the interpretation that arises. Recall that the denotation of the licensing QUD posited to give rise to this utterance is $\lambda d\,.(\{|x{:}\ be\ (blonde, x)|) \geq d\}$.

Input: QUD: How many people of each region are blonde? Utterance: [[POS C]] [1[t_1-many **Scandinavians** $_{F/CT}$] are blonde]] ~C] (reverse *many*-utterance)	INFO STRUC	FOCUS	ANTI DEF	FAITH PROS	DEACC
many(Scandinavians, be & blonde) *~(Scandinavians, be & blonde)* Contrastive element: *Scandinavians* Subject: Specific		***!			*
many(Scandinavians, be & blonde) *~(be & blonde, Scandinavians)* Contrastive element: *blonde* Subject: Specific (regular interpretation)	***!			*	*
□ *many(be & blonde, Scandinavians)* *~(Scandinavians, be & blonde)* Contrastive element: *Scandinavians* Subject: Specific (reverse interpretation)			***		*
many(Scandinavians, be & blonde) *~(Scandinavians, be & blonde)* Contrastive element: *Scandinavians* Subject: Generic		***!			*
many(Scandinavians, be & blonde) *~(be & blonde, Scandinavians)* Contrastive element: *Scandinavians* Subject: Generic	***!				*
many(Scandinavians, be & blonde) *~(be & blonde, Scandinavians)* Contrastive element: *blonde* Subject: Generic (regular interpretation)	***!			*	*
□ *many(be & blonde, Scandinavians)* *~(Scandinavians, be & blonde)* Contrastive element: *Scandinavians* Subject: Generic (reverse interpretation)			***		*
etc.					

Taken together, these optimizations suggest that manipulating or controlling for the interpretive possibilities of the plural subjects of the *many*-utterances used here could have resulted in clearer results, i.e. if the interpretation of stage-level trials and individual-level trials in the Reverse condition were limited to only specific interpretations, reliable differences in response choices and confidence ratings that tracked with those selection choices may have been found, as they were individual-level predicate trials in the Regular condition.

Considering the experimental stimuli, it may be the case that the interpretive possibilities of the subjects are conflated within the stage-level trials, thereby compromising our ability to detect a context-dependent difference in selection rates in these *many*-utterances. Indeed, this seems to be the case considering these example stimuli in (5)-(8):

(5) Many bass hide in shallow areas. *Stage-level, generic interpretation*

(6) Many students publish poetry. *Stage-level, generic interpretation*

(7) Many pundits thought Perlmutter would win. *Stage-level, specific interpretation*

(8) Many economists predicted a raise in interest rates. *Stage-level, specific interp.*

Perhaps with stage-level stimuli balanced for specifically and generically interpreted plural subjects, a clearer interpretive preference between the regular and reverse utterances may have been detected. That is, context may prove to be a strong determinant in the interpretation of *many*-utterances, when a particular interpretation is grammatically licit, based on the interpretation of a plural subject.

To summarize, the preliminary OT analysis has yielded some valuable insight into the regular-reverse ambiguity, demonstrating first and foremost that the interpretive possibilities of a *many*-utterance arise from the interaction with the semantics of the predicate, as well as the information structure and the prosody on the utterance. Further, they suggest the generic-specific ambiguity in plural subjects is another critical factor to manipulate or control for in further experiments. However, other methodological issues may also warrant consideration in further research. For example, in an effort to reduce the complexity of the paragraphs, the structure and contents of the comparison class was not explicitly manipulated (e.g. how many Andorrans or Germans have won the Nobel Prize). This left the full specification of the information structure up to each individual participant. More closely controlling this may reduce some of the unaccounted-for variance in response rates, and enhance the experiment's ability to detect the hypothesized effect of context on selection rates.

However, because this is just a sketch of an OT analysis, several other issued remain to be worked out. For example, the exhaustive list of logically possible output candidates would need to be generated and evaluated against the proposed ranking of constraints here to ensure that only the winners identified here are indeed the optimal candidates. After that, a complete OT-based account of the regular-reverse ambiguity could confirm that the constraints posited here and the ranking assumed, do indeed derive the same interpretive possibilities as those sketched above.

In conclusion, the results of **Experiment 1** and **2** taken together, partially support the main claim that discourse-licensed Focus marking does play a role in disambiguating between the regular and reverse interpretations of *many*-utterances. The regular response selections rates and confidence ratings recorded in **Experiment 1** and **2** did not fully support the hypothesis that the information structure of the discourse context is signaled by emphasis (italics or prosody), and categorically licenses either the regular or reverse interpretation. Instead considering this phenomenon through the theoretical framework of OT, suggested that the generic-specific ambiguity in plural subjects also plays a role. Further investigations of ambiguous, vague quantifiers like *many* should attempt to define the pragmatic factors that determine this generic-specific ambiguity, and align them with those that determine the regular-reverse ambiguity.

So, to return to the question posed in the title of this chapter, neither Focus, nor the prosodic contour thought to signal new information, alone fully determine the regular-reverse ambiguity in *many*. The regular and reverse interpretations of *many*, instead arise out of a complex interaction of semantic, prosodic, and pragmatic constraints on the interpretive possibilities of all the elements in a *many*-utterance.

4. Is the default meaning of *many* relative and proportional?

4.1. Introduction

This chapter describes two experiments designed to demonstrate that a relative, proportional interpretation is the default meaning of *many*; specifically, that the comparison class is a critical part of calculating the truth-values of a *many*-utterance in a given context. To do so, participants were instructed to verify a *many*-utterance with respect to carefully constructed arrays of different colored dots in a speeded judgement task. Following Lidz, et al. (2011) and Pietroski, et al. (2009), these experiments assume that the linguistic truth-conditions for *many*, like *most*, interface with domain-general quantity representations, i.e. the approximate magnitude representations of the Approximate Number System (ANS; Carey, 2009; Dehaene, 1997), and perceptual representations of sets of objects, i.e. parallel individuation processes(Carey, 2009; Halberda, Sires & Feigenson, 2006). These experiments leverage what psychologists know about the ANS and parallel individuation to look for evidence that participants attempt to track all the sets in the context, as the proposed denotation of *many* predicts. First, the nature of this interface and the ANS are briefly described. Then previous work confirming the validity of this assumption is reviewed. Next, each experiment is described in detail. Finally, the importance of these experimental results is related back to the account developed in **Chapter 2**, and the broader literature.

4.1.1. Interfaces with the Approximate Number System

First, let us consider what it would take for a human to understand and use *many* according to the semantic denotation proposed in **Section 2.3**. First, we must consider the nature of the representations that would underpin the meaning of the degree sets over which the critical neutral segment[22] function operates, and then we must consider the nature of the cognitive system that underlies those meanings, to ultimately understand how the truth-conditions of such a semantic denotation are computed in a particular context.

Language is not a completely isolated cognitive system. In order to produce and comprehend the meanings that language attempts to convey, the linguistic system must interface with a multitude of other cognitive systems. For example, quantifiers like *many* and *most* must be understood as comparisons between quantities or magnitudes. Therefore, they offer an ideal illustration of this claim because they may be verified with recourse to some representation of quantity and some process of differentiating the relevant sets in a given context. In the case where the interlocutor does not have the time or ability to count, quantity can be represented as approximate magnitudes, i.e. the representations of the ANS (Carey, 2009; Dehaene, 1997; Feigenson, Dehaene & Spelke, 2004) and different sets can be individuated and tracked by the visual attention system and visual working memory (Carey, 2009).

The experiments described in this chapter leverage what psychologists know about the ANS and the parallel individuation process to look for evidence supporting the relative proportionality of *many* posited in **Chapter 2**. Specifically, the relative aspect of the denotation predicts that the quantity of elements in the quantified set must be

[22] Recall from **Chapter 2** that the notion of 'neutral segment' (Romero 2015b), is likely some kind of central tendency over the whole comparison class (Solt 2009). In this chapter it is operationalized as the mean for clarity and simplicity.

compared to the neutral segment, which is calculated over the quantity of elements in each of the relevant sets in the comparison class – the proportional aspect. Extensive previous research (reviewed by Carey, 2009; Dehaene, 1997; Feigenson, Dehaene & Spelke, 2004) has demonstrated that when two approximate magnitudes are compared in the ANS, the accuracy with which this comparison process can be accomplished depends on the ratio of difference between the two magnitudes; as the ratio approaches 1:1, discriminability – the ease and accuracy with which the two magnitudes can accurately be differentiated – decreases. For example, both 2 vs. 3 and 8 vs. 12 can be discriminated with the same degree of difficulty by educated adults, but both 8 vs. 9 and 16 vs 18 are considerably more difficult, and may even not be discriminable for some adults – meaning that 8 and 9 are perceived as the same quantity (Dehaene, 1997). Moreover, since the denotation proposed in **Chapter 2** posits that a mean and standard deviation must be computed over the whole comparison class, the cardinality of each individual set must be represented and tracked somehow. Research (reviewed by Carey, 2009) has shown that the perceptual system has the capacity to individually track the cardinality of about three sets in parallel, after which accurate magnitude estimation of multiple sets rapidly decreases (Halberda, Sires & Feigenson, 2006).

Given these limitations on comparing magnitudes in the ANS and parallel individuation in of attention and working memory, Lidz, et al. (2011), Pietroski, et al. (2009), and the experiments described in the rest of this chapter use the magnitude discrimination and parallel individuation as tools for investigating the proposed semantic denotations of quantifiers. By rapidly presenting arrays of dots and manipulating the number of different colored sets, these experiments are able to determine the nature of the relative comparison being computed (**ratio-dependent effects**), and how and whether the different sets are used to compute the magnitudes being compared (**set number-dependent effects**).

4.1.2. Experimental validation of the Interface Transparency Thesis: The verification of the quantifier *most* can be supported by ANS representations

The Interface Transparency Thesis (ITT; Lidz, et al., 2011) states that competent speakers of a language are biased towards verifying the truth-values of a given linguistic utterance in a way that is consistent with a "canonical specification of truth-conditions" (p.227) (i.e. a semantic denotation). In a series of experiments described in more detail below, Lidz and colleagues find support for this thesis by showing that when asked to verify *most*-utterances, people seem to rely on enumerating objects in a scene, even when the scene makes a rapid perceptual strategy, like one-to-one matching, extremely salient and easier (e.g. the difference between panel (a) versus (b) and (c) in **Figure 4-1** below). The experiments in this chapter assume the validity of the ITT, and first replicate Lidz, et al. and then extend these findings to a new experimental design to demonstrate that the comparison class is a critical piece of information when verifying a *many*-utterance, as the pragmatic-semantic account of *many* in **Section 2.3** posits.

In an initial experiment, Pietroski, et al. (2009) posited two possible verification strategies for the sentence "Most of the dots are yellow."

(1) Greater-than [#{x: Dots(x) & Yellow(x)}, #{x: Dots(x) & ¬Yellow(x)}]

(2) One-to-one-plus [{x: Dot(x) & Yellow(x)}, {x: Dot(x) & ¬Yellow(x)}]

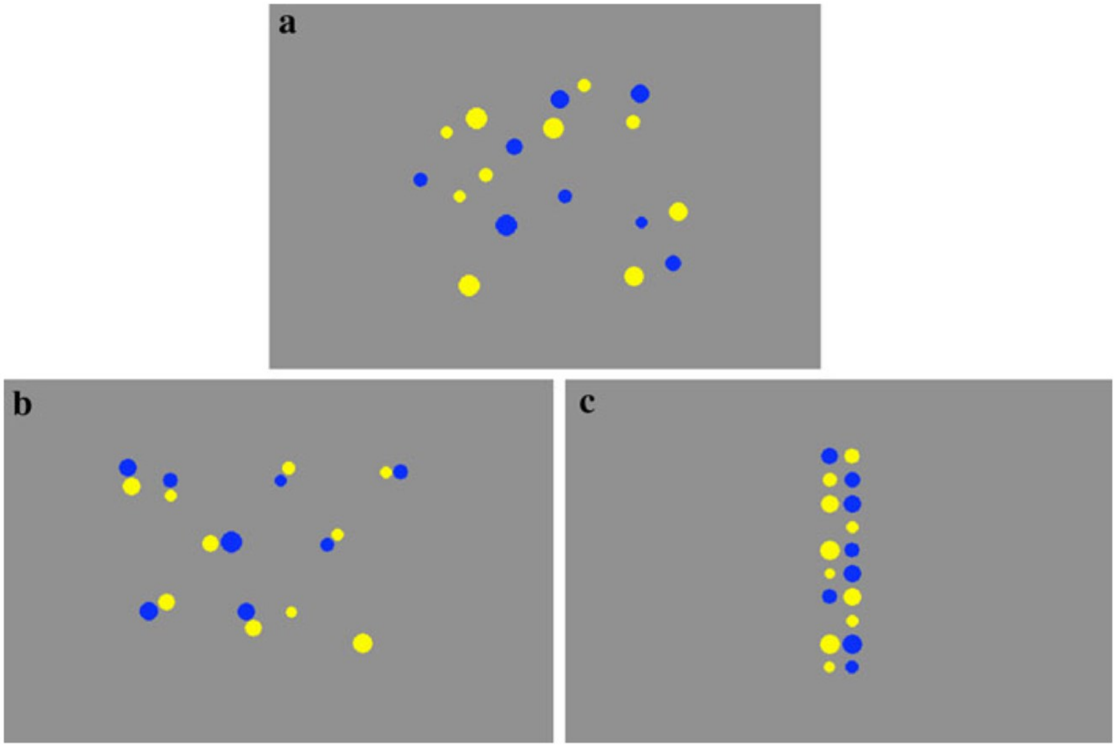


Figure 4-1. Example array types from Pietroski, et al., (2009). (A) Unpaired Scattered; (B) Paired Scattered; (C) Unsorted Column. Participants were shown an array of dots for 200ms and then asked whether the sentence "Most of the dots are yellow," was true or false in relation to the screen they had just seen.

The critical difference between these possible strategies is that (1) requires a count of yellow dots and non-yellow dots, while (2) only requires that the set of yellow dots correspond in a one-to-one relation, to the set of non-yellow dots. While both strategies would result in the same pattern of true and false judgements, Pietroski, et al. (2009) note that only (1) corresponds to the proposed formal semantic truth-conditions of *most* (Hackl, 2009; c.f. Romero 2015b). Moreover, the ANS can only support the processing required by the strategy in (1), because the ANS cannot represent unity (i.e.1), so some other representational system, e.g. an object tracking system, would be required to support the strategy in (2). They presented participants with randomized trials, each containing one of the array types shown in **Figure 4-1**. The array had either random dispersed dots ((a)) or grouped dots ((b) and (c)), in either scattered locations ((c) and (b)) or aligned in columns ((c))[23] and asked them to indicate whether or not the sentence "Most of the dots are yellow," was true or false with respect to the array they had just seen. Note that grouped arrays (e.g. (b) and (c)) in particular lend themselves to the strategy given in (2), because the set of dots that has more items can be easily perceived by just looking for the dots that do not have a pair from the other set. Despite this seemingly more efficient strategy, they found that accuracy decreased as the ratio between yellow and blue dots increased across all three of these array types, suggesting that participants were in fact attempting to discriminate the yellow (Quantified Set) dots from the blue (Comparison Class) dots (a ratio-dependent effect). Moreover, participants could only have pursued this enumeration strategy using their ANS since the rapid presentation (200ms) of the arrays precluded any overt, exact count of the dots. Pietroski et al. thus concluded that participants follow the verification strategy given by a word's semantic truth-conditional denotation.

In a subsequent study, these researchers pursued this issue further asking if the denotation of *most* in fact requires the enumeration of two sets (e.g. the yellow dots and the non-yellow dots) or whether it requires the enumeration of more sets (e.g. each color

[23] In a fourth array type, the dots were sorted by color into vertical columns. With this array type participants' responses were significantly more accurate, even with extremely difficult to discriminate ratios (based on the properties of the ANS). Critically, simulated data derived from a model of the ANS fit the response data from all but this column-sorted trial type suggesting a completely different strategy was employed – one that could not have employed ANS representations (Lidz, et al., 2011). These results offer a caveat to the ITT – the context (i.e. visual display) can override a linguistic specification of meaning in some extreme cases. This point becomes relevant in the **Discussion** of **Experiment 4**.

of dots) (Lidz, et al., 2011). These hypothesized verification strategies are schematized in (3) and (4). Note that (3) is functionally equivalent to the meaning sketched for *most* in **Section 2.3.2** because, unlike (3), it only predicts 'true' responses when the yellow set is the single largest set, requiring participants to individuate in parallel more and more sets as the number of colors in the array increases.[24] Lidz, et al. (like Pietroski, et al., 2009) posited that (3), rather than (4), represents the preferred verification strategy for participants prompted with "Most of the dots are blue."

(3) # of yellow dots > # of all dots - # of yellow dots

(4) # yellow dots > # of red dots + # blue dots

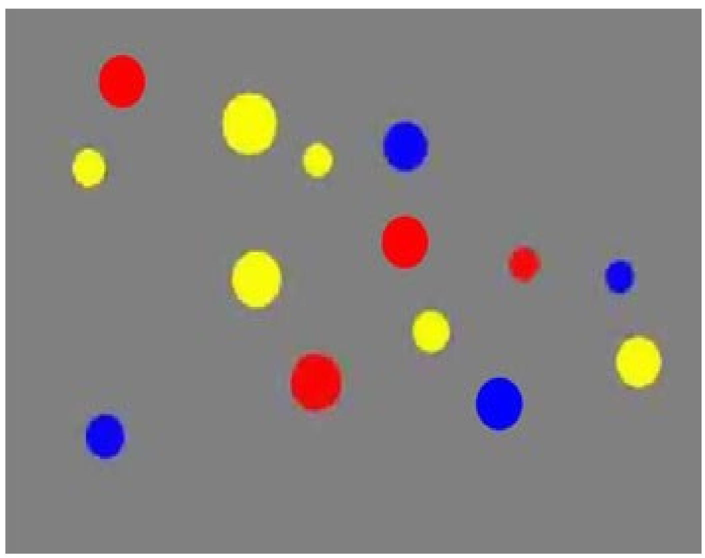


Figure 4-2. An example array from Lidz, et al., (2011). Participants were shown this screen for 150ms and then asked whether the sentence "Most of the dots are yellow," was true or false in relation to the screen they had just seen.

[24] As Tomaszewicz (2013) demonstrates, some languages, e.g. Polish and Bulgarian, have lexically distinct terms for these two meanings. However, these two different *most*s are not clearly lexicalized in English, but can informally be thought of as the difference between *most* (meaning *majority*, relative to the whole) and *the most* (meaning *for the most part*, relative to the other subsets of the whole) (Romero 2015b; Tomaszewicz, 2013).

Lidz, et al. (2011) briefly presented (150 ms) arrays of different colored dots and asked participants to verify whether a *most*-utterance was true with respect to the display. **Figure 4-2** shows an example of their experimental stimuli. They reasoned that if participants follow the verification strategy in (4), accuracy should decrease as the number of sets increases due to the limits on parallel individuation: the adult attentional system and the ANS cannot enumerate more than about three sets in parallel (Halberda, Sires, & Feigenson, 2006), predicting significantly worse performance overall on trials with more than three different colored Sets.

Lidz, et al. (2011) found that participants' ability to correctly accept or reject the sentence relative to the array did not differ as a function of the number of different colored Sets in the array (no set number-dependent effect), suggesting that (3), not (4), is the preferred verification strategy for *most*. They further found that a psychometric model of the ANS, reflecting the ratio-dependent effects inherent in its noisy magnitude representations, closely fit their participants' acceptance patterns. Specifically, participants' accuracy increased as the ratio of yellow dots to non-yellow dots increased (i.e. as the distance between the two magnitudes got larger). Lidz, et al., like Pietroski et al. thus concluded that participants follow the verification strategy given by a word's semantic truth-conditional denotation, and in the case of the quantifier *most*, this requires enumeration that can be supported by the ANS.

In summary, Pietroski, et al. (2009) shows that formal semantic denotations proposed by linguists on the basis of intuited truth-conditions do appear to constrain peoples' default or preferred verification strategies, somewhat independently of whether specific physical properties of context lend themselves to such a strategy or not. Lidz, et al. (2011) further shows that this paradigm can be used to adjudicate between competing hypothesized verification strategies, offering an experimental way to uncover the various ways in which meaning can be specified in the semantic system. Moreover, these results suggest that the non-linguistic ANS can provide magnitude estimates to the semantic system in order to compute the truth-value for a *most*-utterance. In short, these experiments demonstrate how researchers can leverage their understanding of a particular cognitive system, in this case the limits of the ANS and visual attention system, to evaluate linguistic proposals. **Experiment 3** and **Experiment 4** described below attempt to demonstrate that the relative and proportional aspects of the denotation posited in **Section 2.3** do indeed arise as a verification strategy for *many*-utterances, extending and further validating the ITT, and offering experimental support for the account of *many* developed in **Chapter 2**.

Let us now consider a possible verification strategy based on the semantic denotation of *many* proposed in **Section 2.3**. The denotation, repeated here as (5), is relative in the sense that is compares two quantities, and proportional in the sense that the values of those quantities are defined by the quantified set and each of the alternative sets in the comparison class as well; all of which depends on the immediate context[25].

(5) $[\![many_{prop}]\!] = \lambda P_{\langle e,t \rangle}.\lambda Q_{\langle e,t \rangle}.\lambda d_d.(|P \cap Q|:|P|) \geq d$

[25] As opposed to a so-called absolute interpretation (e.g. Partee 1988 summarized in **Chapter 1**) which might compare the Quantified Set to some value not derived from the immediate context.

In the context of the experiments described in this chapter, the relative nature of the denotation proposed in **Section 2.3** suggests that representing the proportions of the main Quantified Set (e.g. the numerosity of dots that are blue; *P* in (5)) and Comparison Class (the numerosity of each set of each color; *Q* in (5)) in a context is critical for calculating the truth-value of a *many*-utterance. Following the reasoning of the ITT, that the semantic denotation does strongly bias competent speakers towards verifying words in a functionally equivalent way to the truth-conditions, then a participant asked to verify a *many*-utterance in a paradigm like Lidz, et al.'s (2011) will need to estimate the magnitude of each alternative set in the Comparison Class and the Quantified Set, and compare the magnitude of the Quantified Set to the mean magnitude of all the sets present in the context. This proposed strategy, schematized as (6) below is in contrast to the much simpler strategy for *most* found by Lidz, et al. (2011), which only requires the magnitude estimation of the quantified and non-quantified sets[26], as schematized in (7) (the verification strategy supported by Lidz, et al.'s results, adapted from (3) above):

(6) Many: $\frac{|QUANTIFIED|}{|ALL|} > \frac{\mu(|ALT1|,|ALT2|,\ldots)}{|ALL|}$

(7) Most: $|QUANTIFIED| > |ALL| - |QUANTIFIED|$

Four illustrative examples are described below, to prime the readers' intuitions about the meanings of these quantifiers, especially the cases in which their truth-values are predicted to differ.

[26] The strategy proposed by Lidz, et al., (2011) was inspired by Hackle's (2009) degree-based account of *most*, and is consistent with Romero's (2015b) findings that the cardinal interpretation of *most* may be the only available interpretation in English.

Referring to the contexts in **Figure 4-3**, consider the truth-values of both (8) and (9) for each context.

(8) Most of the dots are yellow.

(9) Many of the dots are yellow.

Assuming the schematized verification strategies in (6) and (7) above, the predicted verification responses are indicated for each context (true = ✓ and false = #).

First, consider the context in which there are only two different colors of dots: 6 yellow and 5 blue (**Figure 4-3 A**). According to the denotation proposed in **Section 2.3**, *many* is true here. Adopting the denotation proposed by Lidz, et al. (2011; adapted from Hackl, 2009, see also Romero, 2015b) *most* is also true.

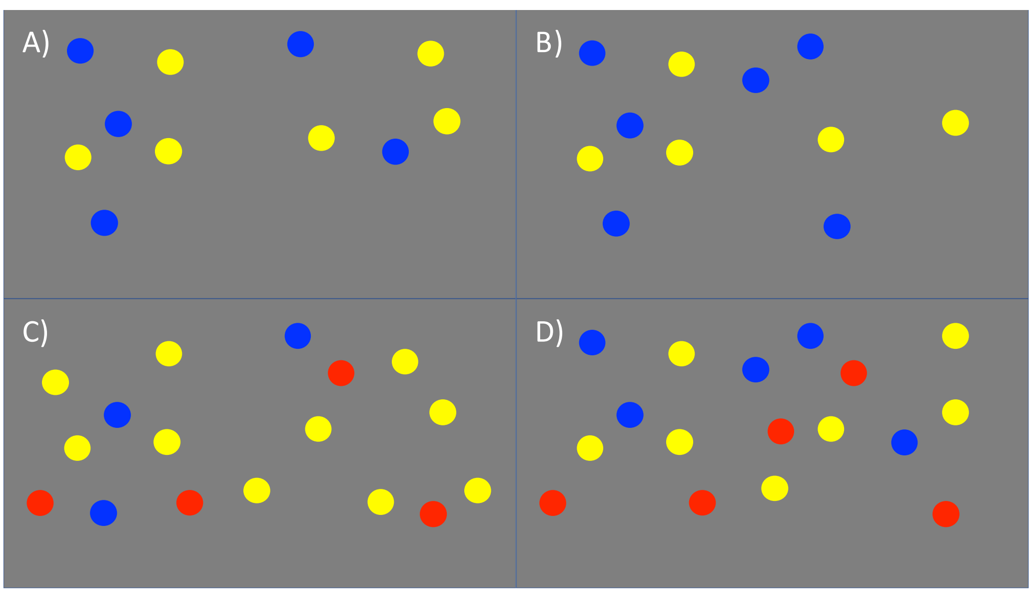


Figure 4-3. Example stimuli to illustrate contexts with just 2 Sets where (a) *many* and *most* are both true, and (b) are both false; and contexts with 3 Sets where (c) *many* and *most* are both true, and (d) *many* but not *most* is true.

✓ $[\![most]\!]$= 6 > (11 – 6)

✓ $[\![many]\!]$= = 6 ⁄ 11 > μ (5 ⁄ 11)

Now consider the context with 5 yellow dots and 6 blue dots (**Figure 4-3 B**). In this case, neither *most* nor *many* would be acceptable.

$[\![most]\!]$= 5 > (11– 5)

$[\![many]\!]$ = 5/ 11 > μ (6 ⁄ 11)

So far, it seems like *most* and *many* are interchangeable. But the real distinction between *most* and *many* appears with contexts with more than two sets. To demonstrate this, consider the minimally more complex situation where there are now three different colors of dots, 10 yellow, 3 blue, and 4 red dots (**Figure 4-3 C**). In this context, both sentences are correct according to the denotations.

✓ $[\![most]\!]$= 10 > (17– 10)

✓ $[\![many]\!]$= 10/ 17 > μ (3 ⁄ 17 + 4/17)

Finally, consider the context with 7 yellow dots, 5 blue and 5 red dots, now *many* is true, but *most* is not, illustrated in **Figure 4-3 D**.

$[\![most]\!]$= 7 > (17– 7)

✓ $[\![many]\!]$= 7/ 17 > μ (5 ⁄ 17 + 5/17)

In summary, the size of the other sets in the context affects the truth-values of *many*, unlike *most* which is only sensitive to the total number of the remaining, unquantified items.

The difference in these two verification strategies predicts that in a paradigm like that used in Lidz, et al. (2011), participants who are asked to verify *many* would show a decrease in accuracy as the number of different sets increased, unlike participants asked to verify *most*. If this predicted difference in response behavior is observed, simply by varying the number of sets and keeping all else equal, this will be taken as support for the relative, proportional denotation of *many*.

4.2. Overview of Experiments

Experiment 3 and **Experiment 4** aim to validate the relative, proportional denotation of *many* proposed in **Section 2.3**. In particular, these two experiments investigate the claim that in order to calculate the truth-conditions and verify a *many*-utterance, one must perform an averaging operation over all the sets present. As described in **Chapters 1** and **2**, this relative proportionality of the denotation is the critical semantic machinery that gives rise to the apparent vagueness of *many*; because proportions must be compared and these magnitudes depend directly on the context: there is no one threshold value between *many* and *not-many*. Moreover, the magnitudes can be represented by the ANS. Thus, the very representations of the magnitudes themselves are noisy and approximate. This all results in the apparent indeterminacy of the truth-conditions, i.e. vagueness, of *many*-utterances.

Experiment 3 begins by replicating the Lidz, et al. (2011) experiment described in the previous section, and then aims to show that participants' verification responses to *many*, unlike *most*, differ depending on the number of different alternative sets present on a given trial. To that end, arrays of dots of varying colors were briefly presented and participants were asked to verify *many*-utterances in relation to those arrays. Asking people to either accept or reject a *many*-utterance as a description of the array, whilst manipulating the ratio value involved in the hypothesized verification strategy and the number of sets, and comparing results between subjects asked to verify *many* and those asked to verify *most* will offer further validity to the proposed denotation for *many* in **Section 2.3**. However, as discussed in more detail in the **Discussion** of **Experiment 3**, the design of the stimuli in **Experiment 3** limited the validity of the conclusions.

To strengthen and further validate the conclusions from **Experiment 3, Experiment 4** strongly biased participants towards a relative strategy by grouping the dots by color and using the partitive construction in the prompt to be verified. It also improved on the design of **Experiment 3**, by manipulating the ratio and set number orthogonally, allowing us to more clearly interpret differences in response curves and compare responses across the *many*- and *most*-conditions directly.

The results of both these experiments demonstrate that people do in fact interpret *many* in a relative, proportional way – showing differences in their acceptance behavior as a function of the relative proportion of the Quantified Set compared to the other Sets in the context – a trial in these experiments. Indeed, evidence was also found for a similarly relative verification strategy when participants were prompted to verify *most* in a context that strongly biased them toward a relative, proportional strategy, supporting the posited critical role of context in the interpretation of ambiguous, vague quantifiers.

4.3. Aim

Experiment 3 aims to find support for the relative, proportional denotation of *many* proposed in **Section 2.3**, which predicts that participants must enumerate the alternative sets in the Comparison Class when verifying *many*. This experiment first replicates findings from Lidz, et al. (2011), where participants are prompted with *most*, and then compares those results to an identical condition where participants are prompted with *many*.

4.4. Background and Hypotheses

Following Lidz, et al.'s (2011) ITT, the denotation of *many* proposed in **Section 2.3** gives rise to the following hypothesized verification strategies and predictions, with respect to Lidz, et al.'s experimental paradigm:

***Hypothesized verification strategy for* most (adapted from Lidz, et al. (2011))**: according to some semantic accounts (c.f. Hackl, 2009; Romero, 2015b; c.f. also Tomaszewicz's (2013) Most1) the denotation of *most* gives rise to a verification strategy, schematized below, which proposes that the cardinality of the Quantified Set is compared to the difference of the cardinality of the Quantified Set and the total cardinality of all the Alternative Sets; if the cardinality of the Quantified Set is at least one greater than this difference, then the utterance should be judged true in that context.

$$|QUANTIFIED| > |ALL| - |QUANTIFIED|$$

***Hypothesized verification strategy for* many**: according to the proposed semantic-pragmatic account in **Section 2.3**, the denotation for *many* gives rise to a verification strategy, schematized below, that considers the proportion of the Quantified Set to every relevant entity in the context, and compares it to the mean proportion of each of the Alternative Sets in the Comparison Class (i.e. to every relevant entity in the context) [27]; if the proportion of the Quantified Set is greater than or equal to the average proportion of all the alternative sets, then the utterance will be judged true in that context.

$$\frac{|QUANTIFIED|}{|ALL|} > \frac{\mu(|ALT1|, |ALT2|, \dots)}{|ALL|}$$

The critical factor being tested in this experiment is whether *many* in fact requires the verifier to enumerate each Alternative Set, rather than simply compare the Quantified Set to the non-quantified sets, or alternatives, like *most* (which requires just enumerating the Quantified Set, and its superset of all relevant entities in the context).

4.5. Method

4.5.1. Participants

Sixty-two undergraduate students (25 males), all with normal or corrected-to-normal vision, participated in this experiment for course credit. One participant's data is not included in the analyses below because of a technical error in saving their data file. This resulted in data for analysis from 30 participants in the *most*-condition (10 males) and 31 participants in the *many*-condition (15 males). This procedure was approved by the Homewood Institutional Review Board.

[27] Based on the consensus in the semantics literature (Cohen 2001; Romero, 2015; Solt 2009; 2015), this includes the Quantified Sets itself (see Cohen 2001 for discussion of why); intuitively, the items in the quantified set are a member of the whole class of those like items. In the account developed in this dissertation, this is captured by the neutral segment function *L(Q)*, which operates over the whole set *Q*, which contains the Quantified Set.

4.5.2. Design and Materials

In order to test the hypothesis that *many*, unlike *most*, requires the verifier to enumerate each Alternative Set, the same experimental design as Lidz, et al. (2011) was used, except for a few small changes. This experiment follows a mixed design with 5 (Ratio Bin: 1:2, 2:3, 3:4, 5:6, 7:8) x 4 (Sets: 2, 3, 4, 5 color sets) within-subjects factors, and quantifier type as a 2-level (*many*, *most*) between-subjects factor. Ratio Bins contained a range of ratio values, calculated as the cardinality of the Quantified Set divided by the summed cardinalities of all the Alterative Sets, centered around the nominal ratio value of the Bin (e.g. the 1:2 Ratio Bin contained ratio values ranging from 0.41 to 0.57 with a mean value of 0.5). Sets refers to the number of sets of different colored dots on a trial. For example, there may be 2 Sets (5 blues dots and 7 yellow dots) or 5 Sets (7 blue dots, 2 red dots, 2 yellow dots, 2 green dots, and 1 purple dot).

To prevent the total area of a set (in pixels) or the size of a dot from being cues to set cardinality, on half of trials the size of each dot varied so that the total number of pixels for all blue dots was equal to the total number of pixels for non-blue dots on the screen (Area-controlled trials). On the other half of trials, the sizes of all dots were randomly varied up to ±35% of the average dot size in a set, such that the size of the average blue dot was made equal to the average non-blue dot (Size-controlled trials). This design yielded 5x4x2=40 unique trial types. Sample trials from different Ratio Bin x Set number levels are shown in **Figure 4-4**.

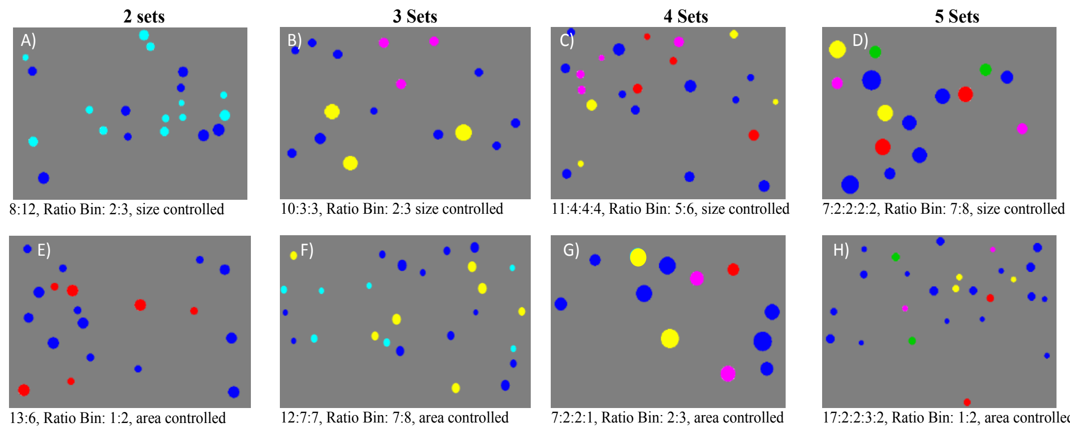


Figure 4-4. Examples of displays generated according to the trial design for **Experiment 3.**

Participants were presented with 10 of each trial-type for a total of 400 actual trials. An additional 12 practice trials were randomly generated from the set of possible trial-types, but were not include in the final data analyzed. Across all trials, the number of possible dots in the Quantified Set ranged from 5 to 17 (with the number of total possible dots on any given trial ranging from 10-32), where the cardinality of the Quantified Set was at least one greater than the cardinality of all Alternative Sets combined. In the other half of the trials, the total cardinality of all the Alternative Set(s) was at least one greater than the cardinality of the Quantified Set (i.e. half of all trials were true, and the other half were false, according to Lidz, et al.'s (2011) verification strategy for *most*).

4.5.3. Procedure

Using the design described above, all trials were randomly generated at the start of each participant's run using the purpose-written Java package from Lidz, et al. (2011). Trials were presented in a unique, random order for each participant. At the start of each trial, the participant was shown the remaining number of trials, and instructed to press the space bar when they were ready to begin the trial. Once they did so, the prompt disappeared, and an array of dots was presented for 150ms. After the dots disappeared, unlike in Lidz, et al., a prompt appeared ("Many/Most dots were blue."[28]) depending on which quantifier condition the participant had been randomly assigned to. Below this prompt were reminders to press the 'J' key if they thought the sentence was true given the array of dots just seen, or 'F' if not. There was no time limit for this response. Once they pressed either key, the next trial would begin. Participants completed the task in roughly 20 minutes on average. **Figure 4-5** shows a schematic of this task flow.

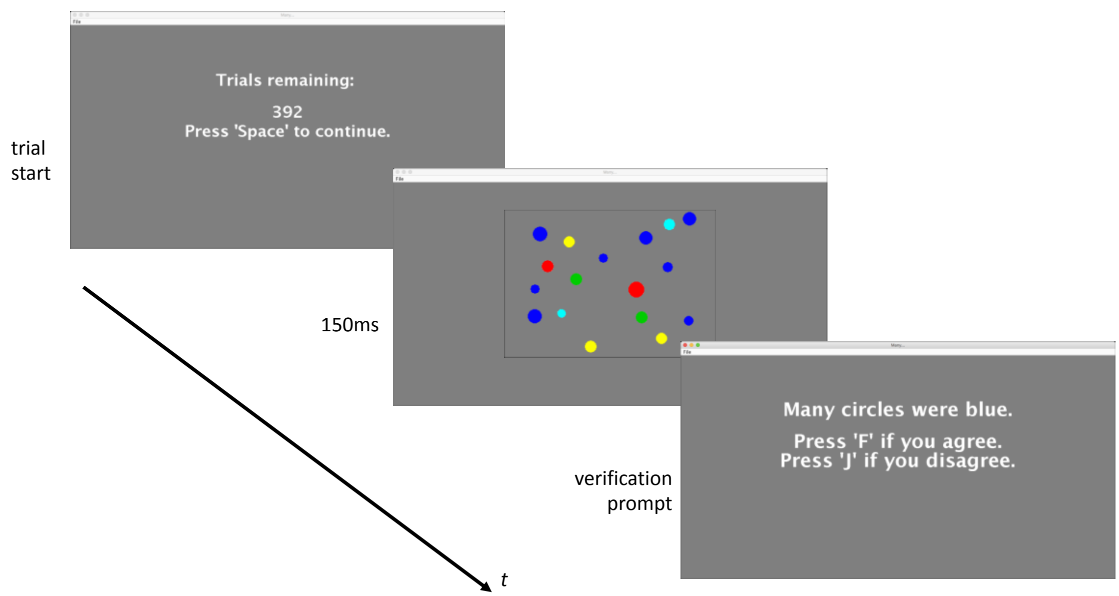


Figure 4-5. A schematic of the screens and task flow used in **Experiment 3.**

4.5.4. Analysis

In Lidz, et al.'s (2011) original experimental design, the ratios in the Ratio Bins were constructed to reflect the hypothesized verification strategy for *most*, so that the

[28] This prompt was used rather than the original "Most/Many of the…", reasoning that a stronger case could be made if evidence of a relative, proportional verification strategy was found in the absence of the overtly partitive construction.

magnitude of the ratio reflected the degree of 'true'-ness or 'false'-ness of *most*. However, that original ratio space does not reflect the hypothesized truth-conditions of *many*, because increases in that ratio magnitude do not necessarily correspond to an increase in 'true'-ness of *many*. For example, if we consider the display in **Figure 4-4 E** with 13 blue dots and 6 red dots, this display sits in very different locations in the two quantifier-specific ratio spaces. For *most*, which compares the Quantified Set to the total cardinality of all the Alternative Sets, the ratio would be calculated as 13/6=2.167. Whereas *many*, by hypothesis, compares the cardinality of the Quantified Set to the mean cardinality of the Sets[29], so the ratio would be calculated as 13/mean(13,6)=13/9.5=1.368. Assuming that the participants are considering each array in the ratio space derived from these verification strategies, the known ratio limit for distinguishing two cardinalities in the ANS, predicts that verifying a "Most…" prompt on this kind of trial would be much easier than verifying a "Many…" prompt. Therefore, to analyze the results of this experiment, two different Ratio scales were calculated:

i) A *most*-Ratio derived from the hypothesized verification strategy of *most*. This ratio reflects the cardinality of the Quantified Set divided by the total cardinality of all the Alternative Set(s)[30], following Lidz, et al. (2011).

---

[29] Because the total cardinality of all the items in the context is known for all trials in this experiment, and these items are not being contrasted with any other kind of item, the proportional aspect of the *many*-verification strategy is set aside for now. The calculation in these contexts with and without the denominator are equivalent:

$$\frac{13}{19} > \mu\left(\frac{6}{9}, \frac{13}{9}\right) = 13 > \mu(6,13)$$

[30] This ratio was calculated as |Blues| / |Alternatives| for True trials (i.e. when |Blues| > |Alternatives|), and 2 – (|Alternatives| / |Blues|) for False trials (i.e. when |Blues| < |Alternatives|), for the same reason as above. This was to achieve roughly symmetric data around the hypothesized threshold with a minimal gap between True and False trials.

ii) A *many*-Ratio derived from the proposed verification strategy of *many*, such that the ratio reflects the cardinality of the Quantified Set over the total number of dots divided by the mean cardinality of all the Alternative Set(s) over the total number of dots[31];

Both of these ratios were then binned in increments of 0.1.

Our aim is to determine whether the rate of acceptance responses differs depending on the number of Sets on a trial, specifically if participants were asked to verify "Many…" but not if they were asked to verify "Most…." To answer this question, sigmoid curves derived from a psychometric beta-binomial model of a participant in this kind of perceptual Yes-No judgement task were fitted to the data, using the python package *psignifit 4*[32] (see Schutt, Harmeling, Macke and Wichmann (2016) for more details on this package). This package estimates the posterior distribution over five parameters that describe a sigmoid curve[33], and confidence intervals (CIs) around each parameter value, using Bayesian inference given a prior over these parameter values derived from a beta-binomial observer model (see Fründ, Haenel & Wichmann (2011) for

---

[31] Like *most,* this ratio was calculated as |Blues| / mean(|alternative 1|, |alternative 2|,…) for True trials (i.e. when |Blues| > mean of the alts), and 2 – (mean of the alts / |Blues|) for False trials (i.e. when |Blues| < mean of the alts).

[32] The package and a wiki are available on the Wichmann Lab's github: https://github.com/wichmann-lab/python-psignifit

[33] Assuming the function in (i), adapted from Fründ et al. (2011):

(i) $$\gamma + (1 - \gamma - \lambda)(1 + \exp\left(\frac{\left(-2\log\left(\frac{1}{\alpha - 1}\right)\right)}{w}\right)(x - \theta)^{-1}$$

Where α (the increments of the stimulus intensity) was always 0.1, based on the bins of the ratio bins.

A fifth parameter is also estimated, η, with values between 0-1, where values close to 0 indicate little to no over-dispersion in the data (a measure of variability). The results of this fitting procedure become less reliable with severely over-dispersed data. However, because all estimated η values from the data were between $1x10^{-5}$, $1x10^{-10}$, eta values are not reported or discussed here.

more details on this model). According to their conventional interpretations in the psychometric function literature (c.f. Kingdom & Prins, 2016), two parameters are critical for evaluating the hypothesized set number effect[34]. Width, w, measures the range of the curve where the velocity – the rate of change – is greatest (has steepest slope). Steeper slopes are taken to indicate more precise or categorical judgements because the range of stimulus intensity between unanimous rejection rates and unanimous acceptance rates is relatively narrow. In the context of this experiment, lower width values are an indication that participants' responses more closely approximate the idealized binary truth-conditions (true, false), suggesting that perceptual, pragmatic, or other cognitive factors are less influential, relative to the semantic meaning itself. Threshold, θ, measures the point where the sigmoid tips from mean 'reject' responses to mean 'accept' responses. In the context of this experiment, the threshold is the lowest ratio value that participants judged 'true' on average (e.g. average acceptance rate > 0.5), given the trial-type and the prompt. This parameter can reflect the difficulty of discrimination, since it reflects the

---

[34] Two other parameters are typically considered:

λ: lapse rate – the shift down in acceptance rate (from 1) for stimuli that should be judged 'true' but are rejected as a result of factors independent of the stimulus itself, e.g. a lapse in attention.

γ: guess rate – the shift up in mean acceptance rate (from 0) for stimuli that should be judged 'false but are judged 'true' as a result of stimulus properties, a guess when a ratio is perceptually difficult to distinguish.

These are nuisance variables in the sense that higher values for the lapse and guess rates would indicate the degree to which the range of average acceptance rates is compressed as a result of perceptual noise or uncontrollable variance, e.g. the participant 'giving up' and making a random response. Little variation in these parameters appears in the data as a function of the Set factor, and therefore they are not discussed at length.

point in the ratio space referred to as subjective equality – or where ratios are perceived as 1:1.

Thus, each of these parameters indicates a different aspect of the participants' response behavior in this kind of Yes-No judgement task.

In the psychometric modeling approach, if the fitted parameter values of the sigmoid curve different datasets are distinct, the researcher can reliably infer that those data were generated by a distinct cognitive or perceptual process. With the fitted parameter estimates in *psignifit 4*, the degree to which one curve is distinct from another is defined by the degree of overlap in each parameter value's CI. In the results presented below, the CIs – estimated via the same Bayesian Inference procedure as the parameter values – indicate the range that contains 95% of all estimated values of each parameter, given the data and the beta-binomial prior distribution over these parameters. Thus, the more overlap between two CI ranges for the same parameter, the less likely those data can be characterized by truly different psychometric functions. To the extent that different parameter values describe different sigmoid curves, these CIs can therefore indicate the degree to which the data underlying each parameter estimate is in fact the result of a distinct psychometric model, i.e. whether the participant is completing the task in a different way. In addition to the interpretation of these CIs, a sigmoid curve to the response data from each participant in each Quantifier condition, for each level of the Set factor was estimated, and entered these values into linear regression models to test for statistical differences between w and θ values across each level of the Set factor, in each

Quantifier condition. The figures that follow present response data and the CI ranges estimated by *psignift4*, and tables summarize the results of these linear regressions.

4.5.5. Predictions

First, because this is a direct replication of Lidz, et al. (2011), no effect of number of Sets is expected in the *most*-condition data. That is, the width w and threshold θ parameter values and their CIs should overlap, suggesting that the sigmoid curves, and therefore the way participants respond in this task, does not depend on the number of Sets on a trial. Instead only an effect of Ratio is expected, such that as the trials move away from 1:1 in *most*-ratio space, acceptability should steadily increase and plateau in the positive direction, and steadily decrease and bottom-out in the negative direction – i.e. the sigmoid curve itself.

Critically, an effect of number of Sets is expected in the *many*-condition data, because the hypothesized verification strategy requires the magnitude of each set to be estimated. That is, the width w parameter values should reliably and significantly increase between trials with 3 or fewer Sets and trials with more than 3 Sets, due to the limits of parallel individuation (Halberda, Sires & Feigenson, 2006). The interpretation of threshold θ values is complicated by the way the data is distributed in the *many*-ratio space, but threshold θ values may increase systematically, reflecting the need for increasing distinctiveness between the magnitudes to be compared (larger ratio values) as the number of Sets increases, due to the additional noise introduced by the increased difficulty of tracking ever more sets (Dehaene, 1997). These predictions are depicted visually in **Figure 4-6.**

If participants do indeed require more distinctive ratios on trials with more Sets when asked to verify "Many…", but not "Most…" – as inferred from the fitted curves as just described – this will be taken as evidence in support of the hypothesized verification strategies for these two different quantifiers, because it would suggest that simply by increasing the number of Alternative Sets, response behavior in the task is altered. The effect of Set number can be intuited by tracing a vertical line from any point on the x-axis in **Figure 4-6**, the points at which that vertical line crosses each of the hypothesized curves would represent trials with the same ratio value. In the case of *most*, the hypothesized curves nearly overlap, and therefore all trials with the same *most*-ratio value are predicted to receive roughly the same mean acceptance rate. In the case of *many*, on the other hand, the hypothesized curves are successively less steep and shifted along the x-axis, depicting the change in width w and threshold θ values dependent on the number of Sets, reflecting the prediction that each in a given ratio bin is predicted to receive systematically lower mean acceptance rates as the number of Sets increase, in other words as the number of Sets increase participants show greater imprecision in their acceptance responses and require larger ratio values to assent to "Many…."

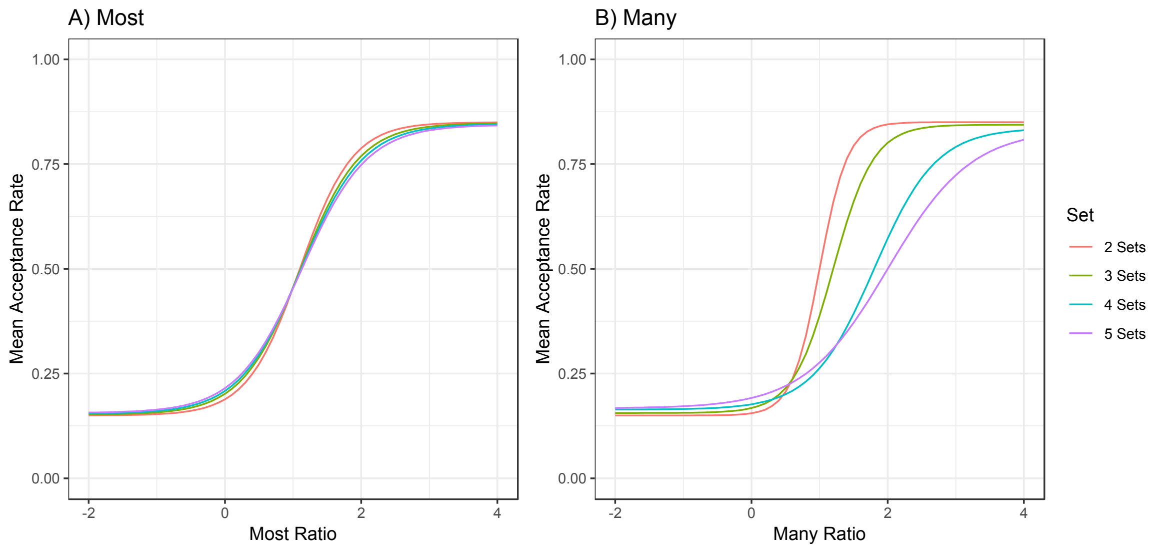


Figure 4-6. This figure illustrates the expected differences in mean acceptance rate as a function of the number of different colored Sets and Ratio (as defined by the verification strategy for each quantifier respectively) in **Experiment 3.**

4.6. Results

The average acceptance rate for each quantifier-specific Ratio bin, across all participants are plotted in **Figure 4-7**. The point in each ratio space at which each quantifier hypothetically shifts from being true to being false is depicted by a red vertical line. Each sigmoid curve was estimated and fit using the procedure described in the **Analysis Section** (**4.5.4**) above, for each Quantifier condition for each different level of the Sets factor. Estimated CIs around each width w (dashed lines) and threshold θ value (solid line) are shown as horizontal lines, corresponding in color to the curve they are estimated from (See **Appendix B** for tables summarizing all the fitted parameter values and CIs). **Table 4-1** summarizes the results of linear regressions for the width w and threshold θ parameter values, with the Set factor coded as the predictor, and participant as a random effect.

First, data from the *most*-condition, as predicted, show no effects of Set number. Referring to the estimated sigmoid curves plotted in **Figure 4-7**, and to the results of the linear regression in **Table 4-1**, note that all the parameter values are co-extensive, and are not significantly different from each other (all *ps* > 0.01). These results suggest that the underlying cognitive processes that give rise to responses do not appreciably differ across trials dependent on the number of different colored Sets. In short, the number of Alternative Sets has no significant effect on the response behavior of participants in this task. These results replicate results reported in Lidz, et al. (2011), validating the method and providing further support for the hypothesized verification strategy for *most*: *most* does not require the verifier to enumerate each Alternative Set separately.

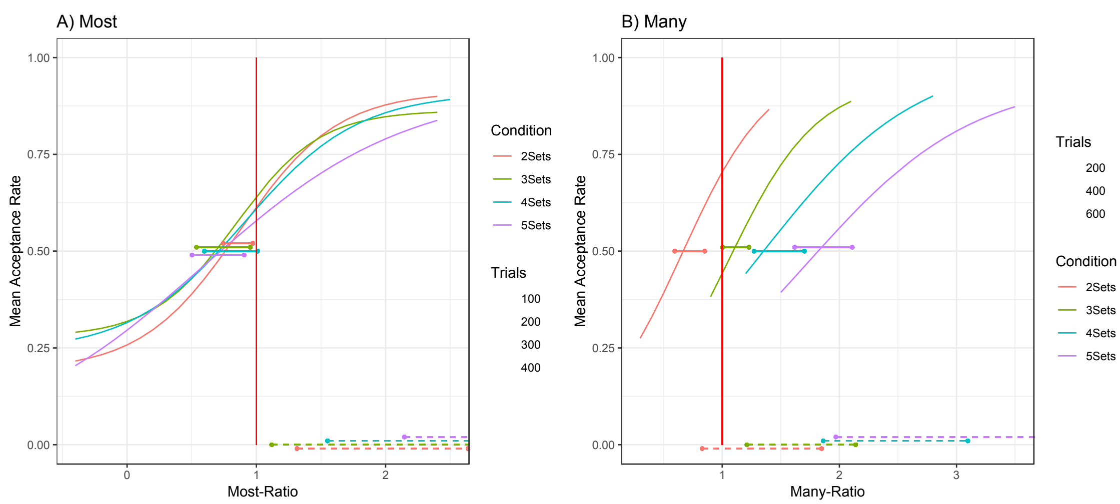


Figure 4-7. Mean accuracy across each Binned Ratio (as derived from the hypothesized truth-conditions for each quantifier) for the (A) *most*- and (B) *many*-conditions in **Experiment 3**. Each color indicates a different number of Sets. Dot size reflects the number of trials in each bin. Dashed horizontal bars represent 95% CIs around the estimated width w value for each fitted curve. Solid horizontal bars represent 95% CIs around the estimated threshold θ value for each fitted curve. The vertical red line represents the point along the ratio axis where the hypothesized verification strategy splits trials into 'true' and 'false'.

Table 4-1. The results of a linear regression with the fitted w and θ values for each participant (respectively) in **Experiment 3,** with the number of Sets as the predictor variable, and participant entered as a random effect.

		w					*theta*			
		Coefficient	se	t-value	p-value		Coefficient	se	t-value	p-value
most	Intercept	2.653	0.267	9.939	0.000		0.666	0.090	7.374	0.000
	2 vs. 3 Sets	0.103	0.377	0.272	0.786		-0.133	0.128	-1.045	0.299
	2 vs. 4 Sets	0.055	0.377	0.145	0.885		-0.011	0.128	-0.085	0.933
	2 vs. 5 Sets	0.091	0.377	0.241	0.810		0.147	0.128	1.150	0.253
	3 vs. 4 Sets	-0.048	0.377	-0.127	0.899		0.123	0.128	0.960	0.339
	3 vs. 5 Sets	-0.012	0.377	-0.031	0.976		0.280	0.128	2.194	0.031
	4 vs. 5 Sets	-0.036	0.377	-0.096	0.924		-0.158	0.128	-1.234	0.220
		w					theta			
		Coefficient	se	t-value	p-value		Coefficient	se	t-value	p-value
many	Intercept	1.321	0.146	9.029	0.000		1.096	0.065	16.879	0.000
	2 vs. 3 Sets	0.010	0.207	0.048	0.962		0.358	0.092	3.901	0.000
	2 vs. 4 Sets	0.460	0.207	2.223	0.028		0.740	0.092	8.062	0.000
	2 vs. 5 Sets	0.665	0.207	3.214	0.002		1.100	0.092	11.984	0.000
	3 vs. 4 Sets	0.450	0.207	2.175	0.032		0.382	0.092	4.161	0.000
	3 vs. 5 Sets	0.655	0.207	3.167	0.002		0.742	0.092	8.083	0.000
	4 vs. 5 Sets	-0.205	0.207	-0.992	0.323		-0.360	0.092	-3.921	0.000

Next, turning to data from the *many*-condition, recall that the hypothesized verification strategy for *many* does require the participant to estimate the cardinality of dots in each Alternative Set(s) in order to compute the mean magnitude of all the sets on the trial. Referring to the curves plotted in **Figure 4-7** and the results in **Table 4-1**, as predicted, there are clear differences in mean acceptance rate as a function the number of different Sets in addition to Ratio, consistent with the hypothesized truth-conditions. Critically, some width w values differed as a function of the number of Sets: specifically, the width w increased between trials with 2 Sets and 5 Sets ($p = .002$), and 3 Sets and 5 Sets ($p = .002$). This pattern of results suggests that, as predicted, when the number of Sets to be operated over increases beyond about three, participants' precision at verifying *many* decreases, compared to trials with just one or two Alternative Sets. In addition, the predicted Set number-dependent difference in threshold θ values was observed in the *many*-condition, unlike in the *most*-condition (all ps < .000), suggesting that verification responses require a successively higher threshold value as the number of Sets increases. However, this difference may be an artefact of the stimuli design as discussed further below.

To summarize, the reliable difference in width and width threshold is indicative of a verification strategy that does require some operation over the Alternative Sets, and serves to differentiate the verification strategies of *many* and *most.*

4.7. Experiment 3: Interim Discussion

Based on the estimated parameter values, their associated CIs and the results of the regressions described above, there are different psychometric models that can be reliably inferred to underlie the observed acceptance rate data from the *many*-condition, and just one from the *most*-condition.

For the *most*-condition, the estimated sigmoid curves suggest that when asked to verify “Most…” participants responded using a strategy that was insensitive to a change in the number of Sets on a given trial. Only the Ratio itself drove differences in acceptance rates in this task. This pattern of results replicates Lidz, et al.’s (2011) results and supports the hypothesized verification strategy for *most*: participants are comparing the cardinality of the Quantified Set to the difference of the Quantified Set and the total cardinality of all the Alternative Sets. They are not performing any operation over each individual Alternative Set.

For the *many*-condition, the estimated sigmoid curves suggest that when asked to verify “Many…” participants responded using a strategy that was sensitive to a change in the number of Sets on a given trial, specifically whether there were 2, 3, or 5 Sets on the trial. The difference in width values suggests that participants' precision decreased as soon as there were more than 3 or 4 Sets on a trial; as the number of Sets increases above about three (the known limit of the perceptual system (Halberda, Sires & Feigenson, 2006)), participants become much less precise in their ability to verify *many*. The difference in threshold values suggests that, as the number of Sets increased, participants required systematically larger Ratios in order to assent to the *many*-utterance. The difference in threshold values could suggest that participants' ability to represent the relevant Ratio may degrade as the number of Sets increases, and so they require larger and larger Ratios to assent. However, as can be seen in **Figure 4-7** and discussed further below, the average ratio value within a given level of the Set factor also increased as the number of Sets increased, so this effect could be an artefact of this property of the stimuli. Despite this, the difference in width values, suggests that participants are attempting to individuate the Alternative Sets, supporting the hypothesized context-sensitive denotation for *many* developed in **Chapter 2**.

Figure 4-8 and **Figure 4-9** summarize these results and their implications. **Figure 4-8** illustrates the average acceptance rate for two trials in the *most*-Ratio space and **Figure 4-9** illustrates the average acceptance rate for two trials in the *many*-Ratio space. These two trials have different *most*-Ratio values, and therefore showed different mean acceptance rates, despite differing in the number of Sets on the trial. Moreover, two trials with the same ratio, but different numbers of Sets (not pictured) also had roughly the same acceptance rates. In contrast, these same two trials have the same *many*-Ratio value (13 blue / mean(13 blue, 6 red)= 11/9.5 ≈ 1.3 and 12 blue / mean(12 blue, 7 yellow, 7 cyan) = 12/16.67 ≈ 1.3), but differ in the number of Sets present (2 vs. 3), and as can be seen from **Figure 4-9**, despite have the same *many*-ratio value, the trial with 3 Sets showed systematically lower acceptance rates compared to the trial with 2 Sets. Trials with the same number of dots (not pictured) also differed in a ratio-dependent way, just

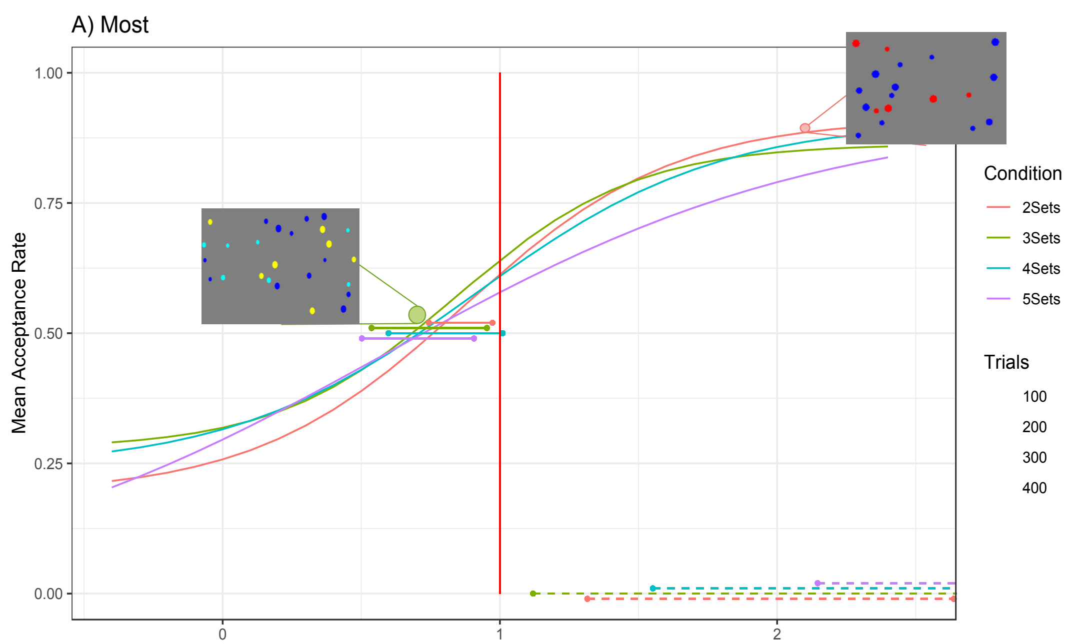


Figure 4-8. Exemplar trials in **Experiment 3** illustrating trials across the full range of *many*-ratio bins and Set numbers, showing the relation between mean acceptance rates and trial-types for the *most*-condition.

like *most*, but this Set number-dependent effect is unique to the *many*-condition. Because the number of dots in each Set, the area of each Set, and the size of each dot is controlled across trials in this experiment, this difference in average acceptance rate can clearly be attributed to the different number of Sets between these two exemplary trials. Thus, because participants across the *most*- and *many*-conditions were given the same kind of stimuli, the fact that differences in acceptance behavior were observed, in a Set number-dependent way, supports the hypothesized verification strategy for *many*, developed from the account in **Chapter 2 Section 2.3**, because it demonstrates that participants who were asked to verify these stimuli relative to a *many*-prompt were sensitive to the Set number manipulation, whereas participants asked to verify a *most*-prompt were not. This sensitivity to the number of Sets is predicted by the verification strategy for *many* but not *most*, because *many* requires the participant to individuate and estimate the magnitude of

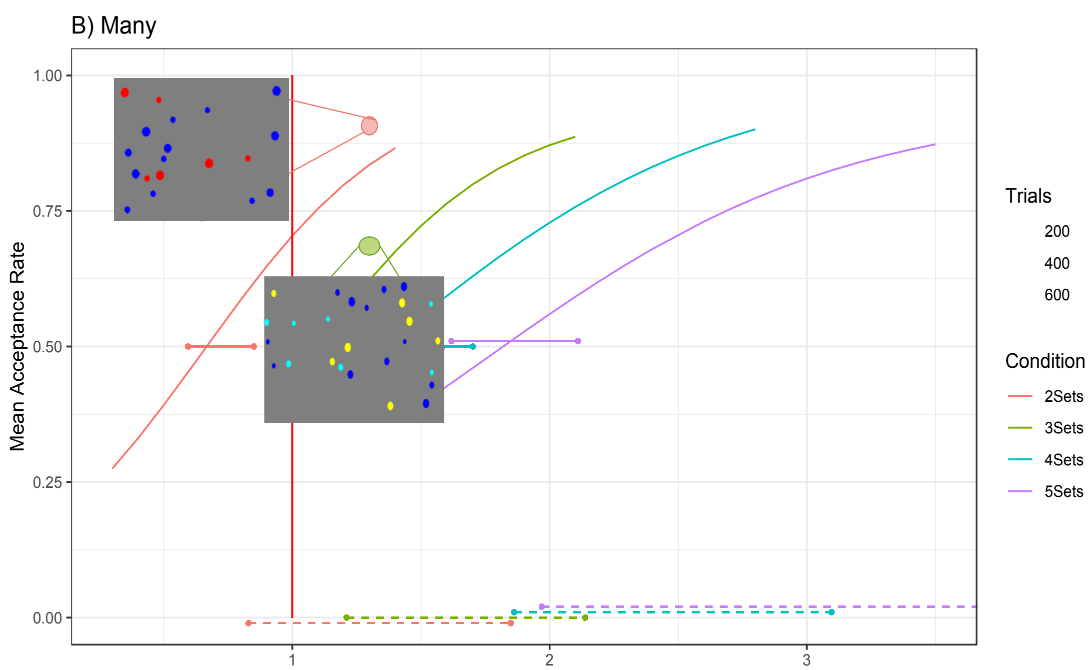


Figure 4-9. Exemplar trials in **Experiment 3** illustrating trials across the full range of *many*-ratio bins and Set numbers, showing the relation between mean acceptance rates and trial-types for the *many*-condition.

each set on a trial in order to compute a mean over all of the Sets' magnitudes on that trial, whereas *most* simply requires participants to compute the difference between the Quantified Set and the rest (i.e. the Alternative Sets do not need to be individuated).

However, as mentioned, certain limitations in the design of this experiment limit the validity of these conclusions. Lidz, et al. (2011) carefully constructed their stimuli so as to sample from the full range of *most*-ratio values within each level of the Sets factor. However, when the data were transformed into the new *many*-Ratio space this aspect of the design no longer holds. This is visually apparent by looking at the way each different curve, and the data from which the curve is estimated, separate along the ratio axes in **Figure 4-9**, but not in **Figure 4-8**. Thus, there is the possibility that the reliable differences in threshold values observed when considering the results in *many*-Ratio space are artefacts of those transformations, rather than real differences in the participants' underlying representations of the meaning of *many*. While the difference in width values (which are less directly tied to the values on the x-axis), argue that there are real differences in the data collected here, **Experiment 4** strengthens and extends our understanding of how people verify *many* and *most* in a rapid perceptual verification task. **Experiment 4** also addresses the lack of data in the tails of the curve, especially in the *many* condition.

4.8. Experiment 4: Aims and Background

Experiment 3 observed Set number-dependent differences in the threshold and width values of sigmoid curves fit to average acceptance rates, when participants were asked to verify a *many*-utterance but not a *most*-utterance. However, these effects may have been an artefact of the experimental design – a confounding of Ratio and Set number in the *many*-ratio space and a lack of data in the extremes of the ratio-space. While Set number-dependent differences in width values in the *many*-condition offer tentative support for the hypothesized relative, proportional meaning of *many*, **Experiment 4** attempts to address the methodological limitations of **Experiment 3** and offer more conclusive evidence of a relative, proportional interpretation of *many*.

Experiment 4 considers the same two quantifiers considered in **Experiment 3** and develops a new design that addresses the limitations of **Experiment 3**. In order to be sure that the differences in threshold and width values are not merely artefacts, the trials in **Experiment 4** are created in such a way that each level of the Set factor spans an overlapping range of Ratio values, in both the *many* and *most* conditions.

The other modification in **Experiment 4** aimed to increase the likelihood of participants adopting a relative, proportional strategy: on each trial, Sets were grouped by color and the overtly partitive construction was used in the prompts (e.g. "Most/Many of the…"). Pietroski, et al. (2009) found that the responses to trials with dots sorted by color into columns as in **Figure 4-10** were not well described by estimated data derived from a psychophysical model of the ANS and were instead better described by responses to a line length estimation task, suggesting that participants were not estimating the numerosity of the dots, but rather comparing the extent of the columns. This demonstrates that in some cases, the visual stimulus clearly engenders a particular strategy. While it remains unclear what, if any, linguistic meaning underlies such a length-based verification strategy, there do exist semantic accounts of a relative *most*

(Tomaszewicz, 2013; Romero, 2015b). Following Pietroski, et al.'s results with the color-sorted columns, and the hypothesized existence of a relative *most*, participants in both the *most*- and *many*-conditions of **Experiment 4** are expected to show Set number-dependent effects, suggesting a relative strategy similar to *many*.

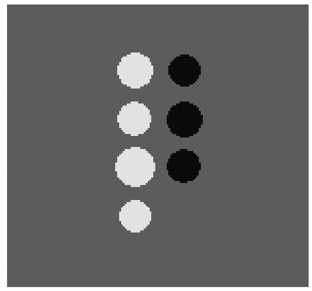


Figure 4-10. Example of Pietroski et al.'s (2009) 'sorted column' trial type. Participants appeared to verify these trials using a length estimation strategy rather than a magnitude comparison strategy.

As the introduction to this chapter briefly describes, Lidz, et al. (2011) propose (10) as an alternative possible verification strategy for *most*.

(10) #yellow dots > # of red dots + # blue dots

(11) # of yellow dots > # of all dots - # of yellow dots

While their participants instead adopted the strategy in (11), the meaning that (10) gives rise to is attested in English and other languages (Romero, 2015b; Tomaszewicz, 2013). In English, the approximate distinction in meaning between (10) and (11) can be paraphrased as *a majority* vs *the most part*, and can surface (modulo certain grammatical constraints, c.f. Pancheva & Tomaszewicz, 2012; Tomaszewicz, 2016) as *most* vs *the most*. Whereas *most* can pick out any majority of relevant items in a context (>51%), *the* *most* (likely in part due to the uniqueness presupposition in the definite article[35]) must refer to the largest subset of items in the context. It is in this sense that the meaning of *the most* may engender a relative strategy akin to *many*, i.e. verifying *the most* requires the hearer to enumerate each Set in the context in order to judge whether the Quantified Set is indeed the largest.

After describing the new methodology of **Experiment 4**, predictions for the *most*- and *many*-conditions are briefly discussed.

4.9. Method

4.9.1. Participants

Sixty-three members of the Johns Hopkins University community with normal or corrected to normal vision participated, and were paid $5 in exchange for participating. These same participants also participated in **Experiment 2**, described in **Chapter 3**, always participating in this experiment first. Thirty-three people (14 males) participated in the *many* condition and 30 people (12 males) participated in the *most* condition. To ensure that participants were native, or near-native (White & Genesee, 1996), speakers of English, participants were screened using the same survey as in **Experiment 3**, after completion of the task. The participation criteria were the same; based on these criteria, three participants' data were excluded from analysis. This procedure was approved by the Homewood Institutional Review Board.

---

[35] Although, the general explanation likely requires such a presupposition be encoded in the quantifier itself since some languages that lexically distinguish these meanings (e.g. Russian) do not have definite articles (Tomaszewicz, 2013; also thank you to Natalia Talmina for native-speaker judgments in Russian).

4.9.2. Design and Materials

This experiment followed a mixed design with Ratio and Set as within-subjects factors, and the verification prompt (e.g. “Many of the …” or “Most of the …”) as a between-subjects factor.

The number of dots in the Quantified Set was either 7, 9, or 11. Using these numerosities, four Ratio Bins, 9:5, 9:6, 9:9, 9:11 (average Quantified Set to average sum of the Alternatives) were created, comprised of trials where the Quantified Set is very large and should be relatively easily perceived as larger than the Alternatives (e.g. 11:6), to trials where that distinction is perceptually harder (9:6), to trials where the Quantified Set is less numerous than other trials, but the ratio of Quantified Set to Alternative Sets is still difficult to discriminate (7:9), or impossible (7:7).

The Set factor manipulated the number of different colored Sets (e.g. 2 Sets as in **Figure 4-11A** or 3 Sets) and the discriminability of the numerosities of the two Alternative Sets. To achieve the latter manipulation, there were two different kinds of 3 Set trials (i.e. trials where there were two Alternative Sets): ‘doubled’ and ‘halved’ 3 Set trials. **Figure 4-11B** shows a ‘doubled’ 3 Set trial (e.g. a 2 Set trial with 9 Blues and 4 Alternatives), and **Figure 4-11C** shows a ‘halved’ 3 Set trial (e.g. 9 Blues, and two Alternative Sets with 2 dots each).

As in **Experiment 3**, low-level visual features of the displays were controlled for. For each Ratio x Set trial type, there was a Size-controlled and an Area-controlled version, resulting in 72 unique Ratio x Set trials.

Finally, the color of the Quantified Set (e.g. whether the prompt stated "Many of the dots were blue," or "… yellow.", or "…red.") was varied across trials, by repeating each of the 72 trial types, for each of the three dot colors (216 trials). This manipulation was intended to force participants to estimate the magnitude of all the Sets present on every trial, rather than fixate on only one color[36]. Twelve of these possible trials were randomly selected and presented as practice trials at the beginning of each run; data from those trials were not analyzed. This resulted in a total of 228 trials.

As **Figure 4-11** illustrates, Sets were grouped by color, such that each Set clustered around a screen-centered fixation cross that remained throughout the whole experiment. Each Set clustered around a center point, equidistant from the central fixation cross, such that all members of the Set were closer to each other than to a member of another Set. It is assumed that these clusters should make a strategy that relies on tracking each Alternative Set even more salient, so that even if there are a variety of possible strategies for verifying *many* or *most*, a strategy that tracks the cardinality of each the Alternative Sets is the most salient and easiest to pursue.

[36] Halberda, Sires and Feigenson (2006) demonstrated that then when target color was prompted before the array was displayed, participants did not display the critical set-number dependent effect.

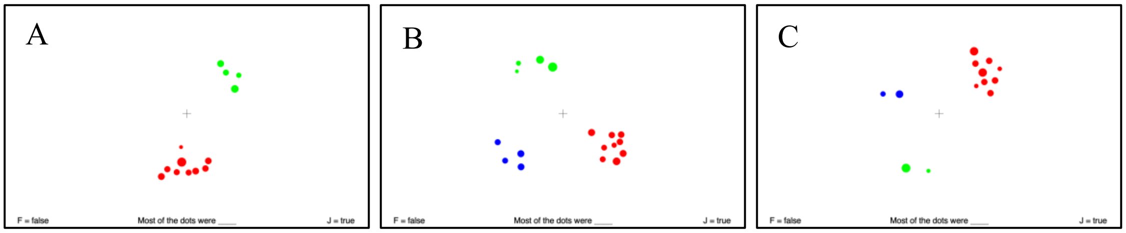


Figure 4-11. Examples of trials for **Experiment 4** where the Quantified Set is blue, (A) 2 Set, size-controlled, 9:4 ratio trial, (B) 'doubled' 3 Set, size-controlled, 9:4:4 ratio trial, and a (C) 'halved' 3 Set, area-controlled, 9:2:2 ratio trial.

4.9.3. Procedure

The procedure for **Experiment 4** is very similar to that of **Experiment 3** except for minor differences described here. Participants were randomly assigned to one of the two between-subjects Quantifier conditions. The prompt that each participant was asked to verify, dependent on the participant's randomly assigned condition, was displayed at the bottom of the screen throughout the experiment. A central fixation cross, as pictured in **Figure 4-11**, also remained in the middle of the screen throughout the experiment. Cues reminding the participant of the response keys (F or J) were also displayed continuously at the bottom of the screen.

On a trial, the array of dots was displayed for 600ms along with the fixation cross, quantifier prompt, and key cues. After the dots disappeared, the target color word was displayed above the fixation cross. Participants were instructed to imagine 'plugging that word into' the prompt sentence at the bottom of the screen, and then instructed to decide whether the sentence was a true description of the dots they just saw. They were given a further five seconds to make their response. After they made a response, or the response time expired, a screen displaying the number of trials remaining appeared. Participants were told they could take brief breaks on this screen, and press the space bar to initiate the next trial.

Practice trials were presented in a random order, but were always the first 12 trials. All actual trials were presented after these, in a pseudo-random order across all participants in all conditions, such that there were never more than three trials with the same target color in a row. On average, participants completed this task in 10 minutes.

4.9.4. Analysis

The response data were analyzed using the same estimated psychometric curves and linear regressions as in **Experiment 3**.

4.9.5. Predictions

Recall the two hypothesized verification strategies for *most* and *many* from **Experiment 3**, repeated in (12) and (13).

(12) Most of the…

$$|QUANTIFIED| > |ALL| - |QUANTIFIED|$$

(13) Many of the …

$$\frac{|QUANTIFIED|}{|ALL|} > \frac{\mu(|ALT1|, |ALT2|, \ldots)}{|ALL|}$$

As in **Experiment 3**, response behavior predicted by the strategy in (12) should be insensitive to the Ratio and Set manipulations implemented in **Experiment 4** because (12) is insensitive to an increase in the number of Alternative Sets. Therefore, the strategy in (12) predicts that the Ratio values will again be the only relevant manipulation in the *most*-condition, and no significant differences in width w or threshold θ parameters across the levels of the Set factor are predicted.

In contrast, the strategy in (13) for *many* suggests that the verifier needs to enumerate the Alternative Sets. So, the strategy in (13) as in **Experiment 3**, predicts the characteristic Ratio effect, and an increase in width w and threshold θ values reflecting the increased difficulty associated with tracking an increased number of Sets. Given the new design of **Experiment 4**, this strategy also predicts an increase in width w and threshold θ values on 'doubled' 3 Set trials, relative to the other two kinds of trials, because of the combined difficulty of representing three Sets (Halberda, Sires & Feigenson, 2006), and also the relative difficulty of representing the comparatively larger, identical Set magnitudes on this kind of trial compared to the 'halved' 3 Set trials (Dehaene, 1997).

Finally, if the new display and prompt style in **Experiment 4** do in fact invite a relative strategy, participants in the *most*-condition may also adopt a relative verification strategy for *most*. To capture the meaning of this relative *most*, the verification strategy in (14) is posited, adapted from semantic accounts of *the most* in the literature (c.f. Romero, 2015b; Tomaszewicz, 2016).

(14) "…most of the…"[37]

$$|QUANTIFIED| > \iota MAX(|ALT1|, |ALT2|, \ldots)$$

[37] Note that this construction cannot take subject position in English, providing clues to certain syntactic facts that may provide an explanation for certain scopal relations that appear to impact the Focus-sensitive interpretation of *most* and *the most*; at first blush, it could be that the uniqueness presupposition blocks the necessary QR that allows F-association. This certainly invites further analysis, perhaps using OT, given the apparent optionality and the hint that syntactic blocking may be at play. This is left for future work.

The verification strategy in (14), similar to (13), suggests that participants must represent the magnitude of each Set in a trial, in order to compute which is the single largest (ιMAX). Therefore, this strategy predicts that, like participants in the *many*-condition, participants in the *most*-condition may show systematically higher width *w* and threshold θ values on trials with 3 Sets compared to just 2. However, this strategy entails a comparison to the largest Alternative Set (via the ιMAX function) rather than the average, meaning that 'doubled' 3 Set trials should not be any harder than a 2 Set trial (modulo the Ratio value), because the two Alternative Sets are always the same magnitude on those trials; this predicts that unlike responses in the *many*-condition, width *w* and threshold θ values will be higher on 'halved' 3 Set trials compared to 2 Set trials and 'doubled' 3 Set trials, reflecting the increased difficulty of tracking 3 Sets compared to 2, and the need to represent two distinct magnitudes instead of just one.

These predictions are visually illustrated **Figure 4-12** (using the estimated parameter values from **Experiment 3** as a basis) and summarized in **Table 4-2**.

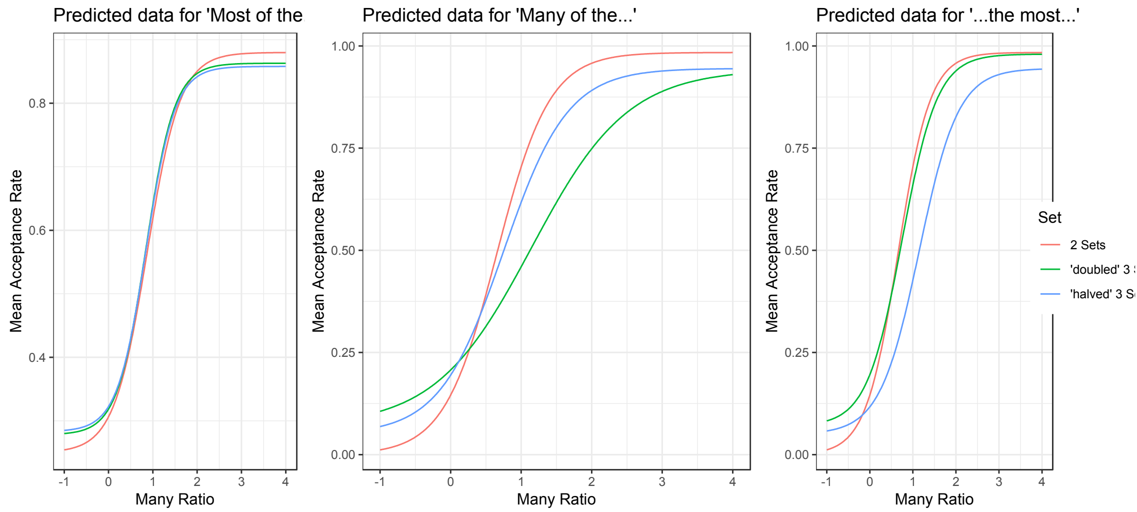


Figure 4-12. Shows predicted sigmoid curves *most*, *many*, and *the most* in **Experiment 4**, based on values from **Experiment 3** and the hypothesized verification strategies described above.

Table 4-2. Predicted parameter values for the three hypothesized verification strategies considered in **Experiment 4.**

	w	*theta*
Most of the…	no differences	
Many of the…	increase in values across Sets, especially with 'doubled' 3Set trials	increase in values across Sets, especially with 'doubled' 3Set trials
The most…	increase in values across Sets, especially with 'halved' 3Set trials	increase in values across Sets, especially with 'halved' 3Set trials

To summarize, the design of **Experiment 4** tests the three hypothesized verification strategies by allowing us to compare changes in mean acceptance rate as a function of manipulating:

(i) the Ratio of the magnitude comparison between the Quantified Set and the Quantifier-specific comparison;

(ii) the number of individual Sets to be tracked;

(iii) the discriminability of the Alternative Sets and the resulting comparison this manipulation gives rise to.

The results of **Experiment 3** are expected to be replicated the *many*-condition; finding that participants asked to verify a *many*-utterance do attempt to enumerate each Alternative Set, and consider the magnitude of the Quantified Set relative to the average magnitude of all the different Sets on the trial. Because the trials in **Experiment 4** all draw from roughly the same range of *many*-Ratio space, the interpretation of these relative differences in response patterns as a function of the Set manipulation will strengthen support for this hypothesized verification strategy for *many*.

Furthermore, because this new task is expected to engender a relative strategy, it is possible that responses in the *most*-condition pattern more like the *many*-condition, as illustrated in the far right panel of **Figure 4-12** instead of the first.

4.10. Results

Initial analysis of the data in the *most*-condition revealed different acceptance rate patterns across the different levels of the Set factor, consistent with the hypothesis that *the most* involves a relative comparison process (i.e. between the Quantified Set and the largest Alternative Set). Therefore, to enable clearer comparison across the two conditions, the data were transformed from the *most*-condition into *many*-Ratio space, as shown in **Figure 4-13**, and consider all results in this common Ratio space (See **Appendix B** for tables summarizing all the fitted parameter values and CIs). **Table 4-3** summarizes the results of linear regressions for each parameter value, with the Set factor coded as the predictor, and participant as a random effect. As in **Experiment 3**, the extent to which the CIs overlap is indicative of the degree of difference in any parameter value, and the results of the linear regression models largely corroborate these CIs.

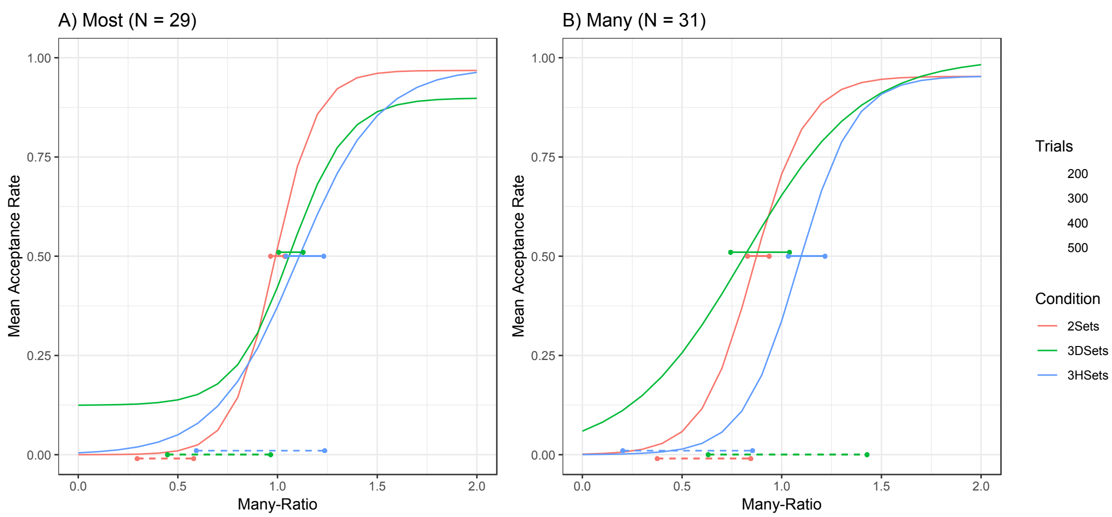


Figure 4-13. Average acceptance rates for each Ratio bin value for each level of Set for data from the (A) *most*-condition, and the (B) *many*-condition in **Experiment 4,** both plotted in *many*-Ratio space. Solid horizontal bars indicate the CI around the estimated threshold values. Dotted horizontal lines represent the CI around the estimated width values.

Table 4-3. The results of a linear regression with the fitted *w* and θ values for each participant (respectively) in **Experiment 4,** with the number of Sets as the predictor variable, and participant entered as a random effect.

		w					*theta*			
		Coefficient	se	t-value	p-value		Coefficient	se	t-value	p-value
most	Intercept	0.403	0.039	10.424	0.000		0.980	0.022	43.574	0.000
	2 vs. 3D Sets	-0.055	0.055	-1.007	0.317		0.054	0.032	1.690	0.095
	2 vs. 3H Sets	0.218	0.055	3.979	0.000		0.280	0.032	8.803	0.000
	3D vs. 3H Sets	-0.273	0.055	-4.986	0.000		0.226	0.032	7.113	0.000
		w					theta			
		Coefficient	se	t-value	p-value		Coefficient	se	t-value	p-value
many	Intercept	0.344	0.053	6.458	0.000		0.894	0.019	47.788	0.000
	2 vs. 3D Sets	0.322	0.075	4.274	0.000		-0.038	0.026	-1.433	0.155
	2 vs. 3H Sets	0.038	0.075	0.503	0.616		0.315	0.026	11.910	0.000
	3D vs. 3H Sets	-0.284	0.075	-3.771	0.000		0.353	0.02647	13.344	0.000

Considering the *most*-condition first, the data clearly support the hypothesized relative strategy. Contrary to the results from **Experiment 3** (and Lidz, et al., 2011;

Pietroski, et al., 2009), but as predicted by the verification strategy posited for *the most*, width w values also differed depending on the number of Sets and their relative magnitudes. As expected, width w values were significantly higher on 'halved' 3 Set trials compared to 2 Set trials ($p < .000$), and there was no difference between 2 Set trials and 'doubled' 3 Set trials ($p = .317$). Unexpectedly, width w values were significantly lower on 'halved' 3 Set trials compared to 'doubled' 3 Set trials ($p < .000$). Threshold θ values also differed significantly as a function of the Set manipulation: critically threshold θ values were significantly higher on 'halved' 3 Set trials compared to 2 Set ($p < .000$), and 'doubled' 3 Set trials ($p < .000$). This pattern of results suggests that when participants were asked to verify "Most of the dots were…," they were least precise and found it systematically harder to discriminate the magnitudes to be compared when there were two Alternative Sets, each with a different magnitude. However, their precision also suffered even if the two Alternative Sets were the same size, but only relative to the hardest 'halved' 3 Set trials and not when compared to 2 Set trials.

Turning to the results for the *many*-condition, the data also partly support the hypothesized relative, proportional verification strategy. As predicted, width w values were significantly lower on 'doubled' 3 Set trials compared to 'halved' 3 Set trials ($p < .000$). Moreover, width w values differed depending on the number of Sets, as expected: width w values were higher on 'doubled' 3 Set trials compared to 2 Set trials ($p < .000$) and lower on 'halved' 3 Set trials compared to 'doubled' 3 Set trials ($p < .000$), and were not significantly different between 2 Set and 'halved' 3 Set trials ($p = .616$). Moreover, threshold θ values differed depending on the number of Sets, but contrary to predictions there was no difference between 2 Set trials and 'doubled' 3 Set trials ($p = .155$). However, there were significantly higher threshold θ values on 'halved 3 Set trials compared to 2 Set trials ($p < .000$). This pattern of results suggests that participants who were asked to verify "Many of the dots were…," found it systematically harder to discriminate the magnitudes to be compared when there were two Alternative Sets, each with a different magnitude. However, their precision only suffered when there were two Alternative Sets and they were the same magnitude.

4.11. Experiment 4: Interim Discussion

To summarize, the results of **Experiment 4** expanded the possible verification strategies available for *most*, demonstrating that in a context with exceedingly salient subsets, a relative interpretation can surface. Furthermore, the results from the *many*-condition confirm that the preferred verification strategy for *many* is sensitive not only to the number of Alternative Sets in the context, but also their relative cardinalities (the structure of the Alternative Sets). The results across the two conditions are strikingly similar, supporting the initial assumption that this new experimental context would engender a relative strategy in the *most*-condition, but show some critical differences supporting the closely related but distinct verification strategies posited for *the most* and *many*, in (13) and (14) above.

This experiment addressed the a conflation of Ratio and Set number in **Experiment 3**, as can be seen by considering the distribution of data points across the three levels of the Set factor in **Figure 4-13**, resulting in stronger conclusions to be the drawn. The ease of discriminating the two magnitudes to be compared similarly depended on a combination of the structure of the Alternative Sets and the number of Alternative Sets, across both Quantifier conditions. However, the precision with which participants were able to judge trials where the magnitude of the Quantified Set was indeed larger than the relevant comparison magnitude (e.g. the average or the largest Alternative Set) differed across the Quantifier conditions. In the *many*-condition, this precision depended on the structure and number of Alternative Sets, whereas this precision depended only on the structure of the Alternative Sets in the *most*-condition.

These results support the hypothesized verification strategies posited above, and offer a more nuanced view of the underlying meaning of *the most* and *many*. Specifically, in the *most*-condition the most difficult trials to assent to were those where there were two Alternative Sets and those Alternative Sets were different sizes. This is consistent with the hypothesized verification strategy if we assume that the algorithm that implements the ιMax function needs to compare the magnitudes of the two Alternative Sets first to identify the larger; since there is no difference between two identical magnitudes this would result less variable performance on 'doubled' trials and trials where this comparison process is superfluous, i.e. 2 Set trials, compared to 'halved' trials (i.e. higher threshold and width values are expected).

In contrast, in the *many*-condition difficultly depended on both the number of Alternative Sets and their magnitudes relative to each other. This is consistent with the hypothesized verification strategy if we assume that the algorithm that implements the *Average* function must individuate and estimate the magnitude of each Alternative Set in turn; when there are two Alternative Sets and their magnitudes differ, the magnitude estimation and subsequent comparison is made easier if the relative difference is greater (i.e. a higher threshold is observed), but even when the two Alternative Sets are the same size, assenting is difficult because estimating the individual magnitudes of two identical Sets may introduce more imprecision due to a more difficult discrimination (i.e. a larger width is observed). In short, assuming this difference in threshold values between 2 'halved' and 3 'doubled' trials is real (and not just the artefact of the distribution of data, which it likely is, given the lack of a reliable difference in the *most*-condition), it could be taken to suggest that the averaging process required by *many* results in even noisier magnitude representations, illustrating another underlying cause of the vagueness of *many*-utterances. This hypothesis requires further investigation to fully understand the correct specification and representation of the neutral segment.

4.12. General Discussion

This chapter described two experiments designed to demonstrate that the default verification strategy for *many* requires the relative comparison of two quantities, and that this comparison can be supported by the ANS. In the results of both **Experiment 3** and **Experiment 4** we found that participants altered their response behaviors depending on the nature of the Comparison Class – how many Alternative Sets were present on a trial and the relative magnitude of those Alternative Sets. **Experiment 4** also demonstrated that the properties of the visual display, and the linguistic signal to be verified can influence the verification strategy that participants adopt. These results broaden the understanding of the interface between linguistic and non-linguistic cognitive representations and offer support for the relative, proportional account of *many* developed in **Section 2.3**.

Experiment 3 found that participants required larger ratios and became less precise as the number of Alternative Sets increased, when asked to verify *many*, but not *most*. These results suggested that the default verification strategy for *many*, but not *most*, was sensitive to the context in a way consistent with a relative, proportional interpretation that requires some representation of the members of the Comparison Class. Then, **Experiment 4** created an environment that aimed to increase the salience of such a relative, proportional strategy, and found that indeed participants asked to verify both *most* and *many* showed sensitivity to the Comparison Class, but in subtly different ways. This finding is reminiscent of participants in Pietroski, et al. (2009), who appeared to verify sorted column trials (**Figure 4-10**) using a length-comparison strategy rather than a magnitude comparison strategy supported via the ANS. In **Experiment 4**, set-based visual appearance similarly invited participants to use a verification strategy distinct from participants asked to verify *most* in **Experiment 3**, where dots were randomly scattered across the screen. These two findings together suggest that verification strategies for quantifiers like *many* and *most* can be represented via the perceptual ANS, and can be influenced by the visual display, and the linguistic input of the prompt. Thus, these data provide support for accounts of ambiguous, vague quantifiers that propose relative comparisons of proportionality, like that in **Chapter 2,** and they provide insight into the situations in which the optimal verification strategy differs from the default truth-conditions, and therefore add nuance to the ITT.

The fact that *many* – and *most*, given the appropriate context – are sensitive to the comparison class in a context, is consistent with the account of *many* (and its extension to *most*) laid out in **Chapter 2** (adopted from Romero, 2015b). Indeed, the conceptual implementation of the verification strategies as magnitude comparison processes supported by the ANS are functionally equivalent to the semantic specification proposed by Romero (2015b) and adopted in **Section 2.3:** according to the semantic denotation, *many* is true when the Quantified Set is greater than or equal to the neutral segment, which can be thought of as the comparison between the Quantified Set and the average distribution derived from the ANS-represented numerosities of all the Sets on a given trial; according to the semantic denotation in **Section 2.3.2** (and Romero 2015b), *most* is true when the Quantified Set is a superset of all the other Sets, which can be thought of as the comparison between the Quantified Set and the largest distribution derived from the ANS-represented numerosities of all the Alternative Sets on a given trial[38].

From a linguistics perspective, the fact that the context can demonstrably shift the default verification strategy, as the set-up in **Experiment 4** did for *most*, provides a clear demonstration of the power that pragmatic inferences – encoded in the contents of the D-tree in the theoretical framework adopted in **Chapter 2** – can have on linguistic comprehension and use. Specifically, the shift from verifying 'Most of the dots are blue' to instead verifying 'The dots are blue, for the most part' could be seen as a kind of context-dependent pragmatic coercion of meaning. In other words, something about the

[38] While this captures the verification strategy adopted for *most* in **Experiment 4,** it does not describe that adopted in **Experiment 3**, suggesting that *most* is also ambiguous, but in a different way from *many*.

context made the latter utterance a more felicitous or salient meaning to verify, despite the surface syntax, giving rise to a different 'default' interpretation (e.g. 'most' not 'the most' appeared in the prompt). In the terminology of **Chapter 2**, the information structure of the preceding discourse licensed a relative interpretation of *most* instead of the majority interpretation, as in **Experiment 3** and Lidz, et al. (2011); here the background could be informed by the prompt, with the content of restriction provided by the visual display (see **Section 2.2.2**). This strengthens the claim that pragmatic factors can play a major role in what possible interpretations arise. However, the current experiments do not, and cannot, provide any insight into the precise nature of the information structure of the discourse that license *the most* instead of *most*, and vice versa.

Future investigations could seek further support for the account in **Section 2.3** by introducing different shapes to test for the effects of Focus investigated in **Chapter 3**, and introduce other manipulations of the linguistic prompt and visual display to understand more about what triggers pragmatic coercion. The experiments in this chapter also leave the existence of the cardinal interpretation, described in **Section 2.3.1** untested. In order to detect and differentiate the cardinal interpretation of *many* from the proportional, the average magnitude of the Quantified Set would need to be manipulated independently of the Set and Ratio manipulations. For example, the mean number of target-color dots across all the trials in **Experiment 4** was about seven, so the cardinal meaning of *many* would predict that trials with less than seven Quantified dots should be judged false and ones with more than 7 should be judged true. There were a small number of trials in **Experiment 4** where the cardinal and proportional interpretations predicted different truth-values, but too few to enable reliable inferences about these responses to be drawn.

In addition, the limits of the perceptual system and how it interacts with linguistic judgments should be further explored. For example, acceptance rates decreased when the number of Sets on the trial exceeded the known limit of the ANS in the *many*-condition, but not the *most*-condition. However, because there were only trials with 2 and 3 Sets, **Experiment 4** did not attempt to replicate this. Perhaps visually separating the sets enables the perceptual system to more accurately track a greater number of sets.

Further, there are yet subtler aspects of the ITT to be explored, and comparisons between the verification strategies explored in these experiments could provide enlightening test cases. For example, as suggested in the **Discussion** of **Experiment 4**, the representations derived via the averaging process in *many* may qualitatively differ from the magnitude estimation processes involved in *most*, or *the most*, i.e. noisier or sharper gaussians. This issue pertains to the correct specification and representation of the neutral segment. The nature of this potential difference could shed light on how different linguistic representations interact with the precision of non-linguistic systems like the perceptual ANS when they are supporting linguistic interpretation. This relative noisiness, and the way it would result in decremented discriminability with ANS representations of magnitude, could also be taken as a one possible implementation of what Kennedy 2019, following Fara 2000, refers to as interest relative properties; essentially, we accept each iterative step of the second premise in the Sorites Paradox because in the classical formulation, the increment of change is so small it results in the judgement that the two degrees are “the same for present purposes” (p. 530).

Indeed, these results of **Experiment 3** and **4** provide support for the relative, proportional interpretation of *many*, and demonstrate the utility of this kind of rapid numerical comparison task for further investigating the meaning of ambiguous, vague quantifiers; informing the best formal mechanisms for capturing their context-dependent shifts in meaning.

5. Conclusion

This dissertation sought to investigate ambiguity and vagueness in language, using the ambiguous, vague quantificational determiner *many* as a test case. Ambiguity can be thought of as many-to-one mapping between multiple meanings and a pronunciation of a lexical item (Kennedy 2011; 2019). Similar but distinct, vagueness can be thought of as the difficulty of determining a clear boundary between true and false (Kennedy 2011; 2019). Ambiguity and vagueness in natural language pose a challenge for formal theories of linguistic meaning because by definition they challenge two core assumptions: (i) that a logically consistent and well-formed theory of meaning must maintain one-to-one mappings from meaning representations to pronounced lexical items (Kennedy 2011; 2019; Parsons 1973); and (ii) that formal meanings can be formulated within a classic two-valued logic with sharp boundaries (Kennedy 2011; 2019).

Previous accounts of *many*, and closely related quantifiers like *few*, *much* and *little* (Cohen 2001; Herburger 1997, 2000; Partee 1988; Romero 2015a, 2015b; Solt 2009, 2011, 2015; Westerståhl 1985; a.o.), have described the various apparent meanings of *many*, and the ways in which its precise range of felicity and truth are difficult to characterize. A recurring idea in this literature is the notion that the particular interpretation of *many* that arises depends, at least in part, on the Focus-marking in a *many*-utterance (Cohen 2001; Herburger 2000; Romero 2015a, 2015b; Westerståhl 1985). However, each of these accounts posits multiple semantic denotations for *many*. Another common idea in this literature, is that the vagueness of *many* can be attributed to a gradable scale encoding either proportions or cardinalities, and/or a comparison between proportions or cardinalities to either proportions or cardinalities on that same gradable scale that are derived from the context, or somehow independently specified (Partee 1988; Romero 2015a, b; Solt 2015).

In **Chapter 2** this dissertation presented its main theoretical proposal. The pragmatic and semantic components used to account for the context sensitive behavior of ambiguous, vague quantifiers was described, including D-trees (Büring 2003; Ginzburg 1997; Roberts 1996/2012), licensing QUDs (Roberts 1996/2012), structured meanings to denote those QUDs (Krifka 2001), information structure-dependent Focus (Rooth 1992), a Focus operator at LF (von Fintel 1994), and a de-compositional, degree-based denotation of *many* (Romero 2015b; Solt 2015). The main claim of this account was that the apparent cardinal-proportional ambiguity, and the regular-reverse ambiguity arise in a principled, predictable way depending on the information structure of the preceding discourse, or context, and the apparent vagueness of *many* arises from the context-sensitive relative proportionality in the denotation of *many* itself. In short, *many* itself is neither ambiguous nor vague, but because it is context-sensitive – and the contexts in which it appears can contain variable, insufficient, or imprecise information – the interpretive possibilities of *many*-utterances can give rise to pragmatic ambiguity and variable truth-value judgements.

Chapter 3 described two experiments designed to test whether participants have access to the reverse interpretation via a particular prosodic contour (c.f. Büring 2003) i.e. whether distinct interpretations arise by virtue of emphasis (taken to mark Focus). To test this, paragraphs were designed that provided discourse contexts with carefully controlled information structure. In **Experiment 1**, participants were asked to summarize the information in the paragraph with either a *many*-utterance with italics consistent with either the regular or reverse interpretations, e.g. "Many Scandinavians have won the *Nobel Prize*," or "Many *Scandinavians* have won the Nobel Prize." They are also asked to rate how confident they were about their choice. No reliable difference was found in the rate at which they chose either kind of *many*-utterance, and confidence rating values only differed when the context licensed the reverse interpretation. However, when the data was re-analyzed with a predicate type factor, participants' choices followed this semantic constraint that suggested the reverse interpretation was ungrammatical on some trials. In **Experiment 2**, a subset of the trials from **Experiment 1** were presented auditorily and the predicate type in the *many*-utterance was balanced across the stimuli. Despite this more careful design, no reliable difference in response choices was found, and again reliable differences in confidence ratings in only the reverse condition. Either the results in **Experiment 1** were spurious, or the null result in **Experiment 2** stemmed from a lack of power. Pending further investigation of these limitations, a preliminary OT analysis (Prince & Smolensky, 1993) considered other semantic constraints on the interpretive possibilities of plural subjects which partly capture the predicate type distinction. This analysis helped illuminate potential factors that could explain the pattern of results across **Experiments 1** and **2**, and suggested that a fuller account of the ambiguity of *many* must incorporate these interpretive constraints on the subjects of *many*-utterances as well.

Chapter 4 described two experiments designed to test whether participants who were asked to verify *many*-utterances took alternatives in the context into account, as the relative, proportional account of *many* predicted. In these experiments, participants were rapidly presented (150-600ms) with different colored dots (2-5 different colors depending on the trial) of varying sizes, either randomly scattered around the screen or grouped by color. They were asked to verify the truth of an utterance with respect to the trial they just saw. In a between-subjects design, half of all participants were asked to verify "Most of the dots were blue," or "Many of the dots were blue." Previous work had posited and found supporting evidence for a meaning of *most* that was not sensitive to the number of alternative sets on the trial. In contrast, the relative comparison in the denotation for *many* predicted differences in performance (acceptance rate) depending on the number of different colored sets present on a given trial; specifically, participants' ability and therefore willingness to accept the utterance would decrease as the number of different colored sets to be tracked increased due to the increased attentional load, thereby causing a change in performance for participants asked to verify *many*, unlike *most*. Across two different experiments, the number of Alternative Sets and their relative magnitude to (i) the Quantified Set (**Experiment 3**), and (ii) to each other (**Experiment 4**) was manipulated. Participants who were asked to verify a *many*-utterance were more sensitive to the number of alternative sets, as predicted. However, when the context strongly biased participants towards a relative strategy, even participants asked to verify a *most*-utterance showed sensitivity to the number of Alternative Sets, suggesting that *most* also has a relative interpretation. Critically, the relative interpretation of *most* seemed to make participants more sensitive to the magnitude of the Alternative Sets relative to each other, but not the Quantified Set, unlike the relative interpretation of *many*, which made participants sensitive to the magnitude of the Alternative Sets relative to each other and the Quantified Set. Thus, these results support the semantic denotation proposed in **Chapter 2**, and demonstrate that a strongly biasing context can determine which of multiple possible interpretations of a lexical item arise, over-riding even the surface syntax.

Thus, the experimental evidence in **Chapters 3** and **4** partially support the theoretical claims made in **Chapters 1** and **2**. Specifically, the fact that a clear, categorical response pattern was not found in **Experiments 1** and **2** suggests that the pragmatic constraints derived from the close relationship between emphasis and information structure (derived from the structured meaning denotations of licensing QUDs in a preceding discourse), are only one part of a complex ecosystem of linguistic constraints that may ultimately be needed to capture the apparent ambiguity of *many*, especially if we want to try and maintain the claim that the regular, reverse, and cardinal interpretations can all be derived from a single relative, proportional gradable denotation. Encouragingly, such a denotation is supported by the results of **Experiments 3** and **4**. In short, *many* is not vague; the boundaries of its truth-conditions are precise and crisp. Rather, *many*-utterances appear vague because they depend on contexts, which may themselves contain unspecified or imprecise quantities, and are interpreted via recourse to noisy, approximate magnitude representations.

The main innovation of the semantic-pragmatic account developed in **Chapter 2** was to use structured meanings to denote the licensing QUDs in discourses. This innovation reduced the abstract theoretical notion of Focus to the actual information structure of the discourse defined by these structured meanings. This approach should be maintained to the extent that it simplifies the notion of Focus, making the contents of the preceding discourse obviously composable with an utterance. Together with semantic constraints on the interpretation of plural subjects, this innovative approach to Focus may prove quite fruitful in accounting for the apparent ambiguity and vagueness of quantifiers like *many*, without having to abandon the assumptions standard semantic theories of reference and truth. Frameworks that facilely allow the modeling of such complex interactions, such as OT (Prince & Smolensky, 1993; or even stochastic or bi-directional OT; Blutner, 2000; Boersma & Hayes, 2001) or Harmonic Grammar (Smolensky & Legendre, 2006), should be employed to further attempt to account for the kind of pragmatic ambiguity displayed by *many*.

Furthermore, the results of the experiments in **Chapter 4** demonstrate the validity of accounts that view vagueness as both a linguistic and meta-linguistic phenomenon (e.g. Fara 2000). Specifically, *many* and *most* are verified in ways consistent with the ratio-dependent magnitude estimation and comparisons processes supported by ANS representations (Dehaene, 1997, Feigenson, Dehaene, & Spelke, 2004; a.o.) as evidenced by the ratio-dependent effects (the sigmoidal shape of the response data). Moreover, *many* but not *most* was also sensitive to the context of a trial, suggesting that the vagueness of *many* likely stems from both the noisy ANS and its context-dependent relative proportionality. This further validation of the ITT (Lidz, et al., 2011) implies that linguistic, lexical representations interface with other conceptual systems during language processing. Therefore, to fully understand and account for the interpretative possibilities in natural language, acceptability judgements must be considered in a rich context, taking into account external facts about the world, and what we know about the way humans perceive and conceptualize that world.

Appendix A: Experiment 1 & 2 Stimuli

This appendix lists the stimuli created for **Experiment 1** in **Chapter 3**. The predicate type is given in parentheses next to the paragraph ID number. Bolded paragraph IDs indicate trials that we reused in **Experiment 2**.

		Regular Condition		Reverse Condition			
		paraphrase order 1	paraphrase order 2	paraphrase order 1	paraphrase order 2	regular response	reverse response
Paragraph 1 (stage)	COCA extract:	Many people thought [Perlmutter] would run for the new 7th Congressional District.					
	Preamble:	All different kinds of election watchers were asked to predict which candidate would win.					
		Of all the pundits asked, two thirds predicted that Perlmutter would win. But one third of all predictions that Perlmutter would win were made by pundits.	One third of all the predictions that Perlmutter would win were made by pundits. But of all the pundits asked, two thirds predicted that Perlmutter would win.	Two thirds of all the predictions that Perlmutter would win were made by pundits. But of all the pundits asked, one third predicted that Perlmutter would win.	Of all the pundits asked, one third predicted that Perlmutter would win. But two thirds of all predictions that Perlmutter would win were made by pundits.	Many pundits thought **Perlmutter** would win.	Many **pundits** thought Perlmutter would win.
Paragraph 2 (stage)	COCA extract:	Adrangi has skewered many Chinese companies listed in the United States.					
	Preamble:	Different hedge fund managers bet against (short) publicly traded companies that are likely to fail for a variety or reasons.					
		Of all the fraudulent companies being traded, two thirds were shorted by Mr. Adrangi. But one third of all companies shorted by Mr. Adrangi were fraudulent.	One third of all companies shorted by Mr. Adrangi were fraudulent. But of all the fraudulent companies being traded, two thirds were shorted by Mr. Adrangi.	Two thirds of all companies shorted by Mr. Adrangi were fraudulent. But of all the fraudulent companies being traded, one third were shorted by Mr. Adrangi.	Of all the fraudulent companies being traded, one third were shorted by Mr. Adrangi. But two thirds of all companies shorted by Mr. Adrangi were fraudulent.	Many fraudulent companies have been shorted by **Mr. Adrangi**.	Many **fraudulent** companies have been shorted by Mr. Adrangi.
Paragraph 3 (stage)	COCA extract:	Since impoundment, these areas have been reduced to woody snarls of downed limbs and standing stumps that line the old channel, creating excellent ambush points where many bass can take refuge in a small area.					
	Preamble:	In Lake Champlain, there a variety of fish species that each have their preferred kind of hiding place.					
		Of all the bass in the lake, two thirds prefer to hide in shallow areas. But one third of shallow areas hide bass at any given time.	One third of shallow areas hide bass at any given time. But of all the bass in the lake, two thirds prefer to hide in shallow areas.	Two thirds of shallow areas hide bass at any given time. But of all the bass in the lake, one third prefer to hide in shallow areas.	Of all the bass in the lake, one third prefer to hide in shallow areas. But two thirds of shallow areas hide bass at any given time.	Many bass hide in **shallow** areas.	Many **bass** hide in shallow areas.
Paragraph 4 (individual)	COCA extract:	In many states where the species once thrived, populations are so low in number and so isolated that they contain a shortage of genetic information.					
	Preamble:	The conservation status of a species depends on a variety of factors related to the species' ability to reproduce viable offspring, or fitness.					
		Of all the species classified as endangered, two thirds lack genetic fitness. But one third of all species that lack genetic fitness are classified as endangered.	One third of all species that lack genetic fitness are classified as endangered. But of all the species classified as endangered, two thirds lack genetic fitness.	Two thirds of all species that lack genetic fitness are classified as endangered. But of all the species classified as endangered, one third lack genetic fitness.	Of all the species classified as endangered, one third lack genetic fitness. But two thirds of all species that lack genetic fitness are classified as endangered.	Many endangered species lack **genetic** fitness.	Many **endangered** species lack genetic fitness.
Paragraph 5 (stage)	COCA extract:	[Tour de France fans] are knowledgeable and opinionated and many of them are drunk, even thought the real talent won't be here for another 26 hours.					
	Preamble:	Different kinds of sports fans find different ways to revel while watching their sport of choice.					
		Of all the Tour de France fans at the race, two thirds of them get drunk. But one third of all sports fans that get drunk at their venue are Tour de France fans.	One third of all sports fans that get drunk at their venue are Tour de France fans. But of all the Tour de France fans at the race, two thirds of them get drunk.	Two thirds of all sports fans that get drunk at their venue are Tour de France fans. But of all the Tour de France fans at the race, one third of them get drunk.	Of all the Tour de France fans at the race, one third of them get drunk. But two thirds of all sports fans that get drunk at their venue are Tour de France fans.	Many Tour de France fans get **drunk**.	Many **Tour de France** fans get drunk.

		Regular Condition		Reverse Condition			
		paraphrase order 1	paraphrase order 2	paraphrase order 1	paraphrase order 2	regular response	reverse response
Paragraph 6 (individual)	COCA extract:	Flowers are becoming more profuse, and planted heliotrope trees have grown in and now define many [golf]					
	Preamble:	All different kinds of native trees are used to define the various features on Hawaiian golf courses.					
		Of all the holes on the course, nine tenths are defined by heliotrope trees. But one tenth of all the features that heliotrope trees define are holes.	One tenth of all the features that heliotrope trees define are holes. But of all the holes on the course, nine tenths are defined by heliotrope trees.	Nine tenths of all the features that heliotrope trees define are holes. But of all the holes on the course, one tenth are defined by heliotrope trees.	Of all the holes on the course, one tenth are defined by heliotrope trees. But nine tenths of all the features that heliotrope trees define are holes.	Many holes are defined by **heliotrope** trees.	Many **holes** are defined by heliotrope trees.
Paragraph 7 (stage)	COCA extract:	Also, indecision at the top has allowed many factory managers to cling to hopes that military budget cuts will					
	Preamble:	In Post-Soviet Russia, all different kinds of manufacturing sectors were having to look for new and different sources of capital.					
		Of all the weapons manufacturers, nine tenths still received government funding. But one tenth of all the manufacturers that still received government funding were weapons manufacturers.	One tenth of all the manufacturers that still receive government funding were weapons manufacturers. But of all the weapons manufacturers, nine tenths still received government funding.	Nine tenths of all the manufacturers that still receive government funding were weapons manufacturers. But of all the weapons manufacturers, one tenth still received government funding.	Of all the weapons manufacturers, one tenth still received government funding. But nine tenths of all the manufacturers that still received government funding were weapons manufacturers.	Many weapons manufacturers received **government** funding**.**	Many **weapons** manufacturers received government funding.
Paragraph 8 (stage)	COCA extract:	And many of these [people published in Poetry Magazine] are still in school, a lot of college and graduate					
	Preamble:	All different kinds of people who write get all different kinds of literary work published.					
		Of all the college students who write poetry, nine tenths get published. But one tenth of all the poetry published is written by college students.	One tenth of all the poetry published is written by college students. But of all the college students who write poetry, nine tenths get published.	Nine tenths of all the poetry published is written by college students. But of all the college students who write poetry, one tenth get published.	Of all the college students who write poetry, one tenth get published. But nine tenths of all the poetry published is written by college students.	Many students publish **poetry**.	Many **students** publish poetry.
Paragraph 9 (individual)	COCA extract:	Marshall Sahlins presented his differentiation of political types, … in 1961, when the highlands of New Guinea					
	Preamble:	In 1960, Westerners were in varying degrees of contact with communities of all sizes in the Papua New Guinean Highlands.					
		Of all the small communities in the Highlands, nine tenths were still isolated. But one tenth of all the isolated communities were small6	One tenth of all the isolated communities were small. But of all the small communities in the Highlands, nine tenths were isolated.	Nine tenths of all the isolated communities were small. But of all the small communities in the Highlands, one tenth were isolated.	Of all the small communities in the Highlands, one tenth were still isolated. But nine tenths of all the isolated communities were small.	Many small communities were **isolated**.	Many **small** communities were isolated.
Paragraph 10 (stage)	COCA extract:	Had Asia not come along, many analysts are convinced that the Fed would have already begun to raise short-					
	Preamble:	In 2011, several different kinds of experts were asked to predict how the Federal Reserve Bank would adjust short-term interest rates.					
		Of all the economists asked for predictions, nine tenths predicted a raise in interest rates. But one tenth of all those who predicted a raise in interest rates were economists.	One tenth of all those who predicted a raise in interest rates were economists. But of all the economists asked for predictions, nine tenths predicted a raise in interest rates.	Nine tenths of all those who predicted a raise in interest rates were economists. But of all the economists asked for predictions, one tenth predicted a raise in interest rates.	Of all the economists asked for predictions, one tenth predicted a raise in interest rates. But nine tenths of all those who predicted a raise in interest rates were economists.	Many economists predicted a **raise** in rates.	Many **economists** predicted a raise in rates.

		Regular Condition		Reverse Condition			
		paraphrase order 1	paraphrase order 2	paraphrase order 1	paraphrase order 2	regular response	reverse response
Paragraph 11 (stage)	COCA extract:	In Kosrae, 90 percent of adult surgical admissions are linked to diabetes, and of these many are for amputations necessitated by vascular breakdown.					
	Preamble:	On the Polynesian island of Kosrae, a high rate of invasive medical interventions result form a range of serious chronic illnesses.					
		Of all the surgeries performed, three fourths are the result of diabetes. But one fourth of diabetes cases result in surgery.	One fourth of diabetes cases result in surgery. But of all the surgeries performed, three fourths are the result of diabetes.	Three fourths of diabetes cases result in surgery. But of all the surgeries performed, one fourth are the result of diabetes.	Of all the surgeries performed, one fourth are the result of diabetes. But three fourths of diabetes cases result in surgery.	Many surgeries result from **diabetes**.	Many **surgeries** result from diabetes.
Paragraph 12 (stage)	COCA extract:	Most council members have jobs, but many attend neighborhood meetings and get calls late at night about					
	Preamble:	In small cities like Berne, Indiana all the different public officials attend a wide variety of public meetings.					
		Of all the city council members, three fourths attend neighborhood meetings. But one fourth of neighborhood meeting attendees are city council members.	One fourth of neighborhood meeting attendees are city council members. But of all the city council members, three fourths attend neighborhood meetings.	Three fourths of neighborhood meeting attendees are city council members. But of all the city council members, one fourth attend neighborhood meetings.	Of all the city council members, one fourth attend neighborhood meetings. But three fourths of neighborhood meeting attendees are city council members.	Many councilors attend **neighborhood** meetings.	Many **councilors** attend neighborhood meetings.
Paragraph 13 (individual)	COCA extract:	Although there as an extensive system of roads from Roman times that was kept in fairly good repair, many					
	Preamble:	Across 5th century Europe the quality of roads varied greatly.					
		Of all the roads in Britain, three fourths were mere trials. But one fourth of all trails were in Britain.	One fourth of all trails were in Britain. But of all the roads In Britain, three fourths were mere trails.	Three fourths of all trails were in Britain. But of all the roads in Britain, one fourth were mere trails.	Of all the roads in Britain, one fourth were mere trails. But three fourths of all trails were in Britain.	Many roads in Britain were mere **trails.**	Many roads in **Britain** were mere trails.
Paragraph 14 (individual)	COCA extract:	That stop will follow Beijing and will accompany stops to Hong Kong: China's ancient capital, Xi'an; and a destination many Chinese call the most beautiful place on Earth: Guilin.					
	Preamble:	Tourists from around the world were asked which place they thought was the most beautiful place on Earth.					
		Of all the Chinese tourists, three fourths voted for Guilin. But one fourth of tourists who voted for Guilin were Chinese.	One fourth of tourists who voted for Guilin were Chinese. But of all the Chinese tourists, three fourths voted for Guilin.	Three fourths of tourists who voted for Guilin were Chinese. But of all the Chinese tourists, one fourth voted for Guilin.	Of all the Chinese tourists, one fourth voted for Guilin. But three fourths of tourists who voted for Guilin were Chinese.	Many Chinese tourists think **Guilin** is the most beautiful place on Earth.	Many **Chinese** tourists think Guilin is the most beautiful place on Earth.
Paragraph 15 (stage)	COCA extract:	In his years in power, Mr. Bonanno shunned the flamboyant styles favored by many contemporary mob bosses, including Charles (Lucky) Luciano, Thomas (Three Finger Brown) Lucchese and Frank Costello, who delighted in wearing elegant clothes and being the hosts of lavish parties in nightclubs in Manhattan and Miami Beach.					
	Preamble:	In the 90s, all kinds of wealthy people were living a range of lifestyles.					
		Of all the mobsters, three fourths preferred a flamboyant lifestyle. But one fourth of wealthy people who preferred a flamboyant lifestyle were mobsters.	One fourth of wealthy people who preferred a flamboyant lifestyle were mobsters. But of all the mobsters, three fourths preferred a flamboyant lifestyle.	Three fourths of wealthy people who preferred a flamboyant lifestyle were mobsters. But of all the mobsters, one fourth preferred a flamboyant lifestyle.	Of all the mobsters, one fourth preferred a flamboyant lifestyle. But three fourths of wealthy people who preferred a flamboyant lifestyle were mobsters.	Many mobsters preferred a **flamboyant** lifestyle.	Many **mobsters** preferred a flamboyant lifestyle.
Paragraph 16 (individual)	COCA extract:	Individuals who are legally blind or visually impaired in the United States have long suffered high					
	Preamble:	Job training programs for a variety of special needs populations exist in varying amounts across the developed world.					
		Of all the job training programs that exist for blind people, four fifths are in America. But one fifth of all job training programs in America are for blind people.	One fifth of all job training programs in America are for blind people. But of all the job training programs that exist for blind people, four fifths are in America.	Four fifths of all job training programs in America are for blind people. But of all the job training programs that exist for blind people, one fifth are in America.	Of all the job training programs that exist for blind people, one fifth are in America. But four fifths of all job training programs in America are for blind people.	Many training programs for blind people exist in **America**.	Many training programs for **blind** people exist in America.

		Regular Condition		Reverse Condition			
		paraphrase order 1	paraphrase order 2	paraphrase order 1	paraphrase order 2	regular response	reverse response
Paragraph 17 (stage)	COCA extract:	Already many see their leaders as kowtowing to the US by fighting militants at its behest and allowing drone attacks on Pakistani territory.					
	Preamble:	A range of stakeholders interested in Pakistan are frustrated with the ineptitude and corruption that plague their different civic institutions.					
		Of all Pakistani citizens, four fifths are frustrated with their government. But one fifth of the people who are frustrated with the Pakistani government are Pakistani citizens.	One fifth of people who are frustrated with the Pakistani government are Pakistani citizens. But of all Pakistani citizens, four fifths are frustrated with their government.	Four fifths of people who are frustrated with the Pakistani government are Pakistani citizens. But of all Pakistani citizens, one fifth are frustrated with their government.	Of all Pakistani citizens, one fifth are frustrated with their government. But two fifths of the people who are frustrated with the Pakistani government are Pakistani citizens.	Many Pakistanis are frustrated with their **government**.	Many **Pakistanis** are frustrated with their government.
Paragraph 18 (stage)	COCA extract:	Many complications are becoming so routine that nurse midwives can manage them instead of sending women to high-risk care specialists.					
	Preamble:	According to a 1980 report on childbirth procedures, a wide range of medical issues were managed by a variety of different healthcare professionals.					
		Of all the complications that arose during childbirth, four fifths were reportedly managed by midwives. But one fifth of midwives reported managing complications.	One fifth of midwives reported managing complications. But of all the complications that arose during childbirth, four fifths were reportedly by midwives.	Four fifths of midwives reported managing complications. But of all the complications that arose during childbirth, one fifth were reportedly by midwives.	Of all the complications that arose during childbirth, one fifth were reportedly managed by midwives. But four fifths of midwives reported managing complications.	Many complications were managed by **midwives.**	Many **complications** were managed by midwives.
Paragraph 19 (stage)	COCA extract:	Many activists and politicians are calling for more aggressive policy now and many in Washington are openly angry with Obama's current policy to stay friendly with both sides of the conflict.					
	Preamble:	During secession talks between North and South Sudan, various stakeholders called for different kinds of policies from the Obama administration.					
		Of all the activists involved, four fifths called for an aggressive policy. But one fifth of stakeholders who called for an aggressive policy were activists.	One fifth of stakeholders who called for an aggressive policy were activists. But of all the activists involved, four fifths called for an aggressive policy.	Four fifths of stakeholders who called for an aggressive policy were activists. But of all the activists involved, one fifth called for an aggressive policy.	Of all the activists involved, one fifth called for an aggressive policy. But four fifths of stakeholders who called for an aggressive policy were activists.	Many activists called for an **aggressive** policy**.**	Many **activists** called for an aggressive policy.
Paragraph 20 (stage)	COCA extract:	Evolution, systematics, and taxonomy have long been well served by NHCs, but ecology, biodiversity, conservation biology, and many other aspects of biological science could also benefit from greater access to the collections' data.					
	Preamble:	In the near future, a range of scientific fields will benefit from an ever growing array of data-related innovations.					
		Of all the sub-fields of biological science, four fifths will benefit from open-access data. But one fifth of all the scientific fields that will benefit from open-access data are biological.	One fifth of all the scientific fields that will benefit from open-access data are biological. But of all the sub-fields of biological science, four fifths will benefit from open-access data.	Four fifths of all the scientific fields that will benefit from open-access data are biological. But of all the sub-fields of biological science, one fifth will benefit from open-access data.	Of all the sub-fields of biological science, one fifth will benefit from open-access data. But four fifths of all the scientific fields that will benefit from open-access data are biological.	Many biological sciences will benefit from **open-access** data**.**	Many **biological** sciences will benefit from open-access data**.**

Appendix B: Parameter values for Experiment 3 and 4 data

The tables in this appendix give the full set of estimated parameter values and their CIs for the sigmoid curves fitted by *psignifit* to the response data from **Experiments 3** and **4.** Values for the following parameters that characterize the best fit sigmoid function for the *many*-Ratio and *most*-Ratio: theta – threshold; w – width; lambda – lapse rate; gamma – guess rate. CIs indicate the range that contains 95% of simulated parameter values following Bayesian inference, given the data and a beta-binomial model-based prior over these parameters; thus, the more overlap between two CI ranges for the same parameter, the less likely those data can be characterized by truly different psychometric functions.

Experiment 3

		w			*theta*			*lambda*			*gamma*			*eta*		
			95% CI			95% CI			95% CI			95% CI			95% CI	
			lower	upper		lower	upper		lower	upper		lower	upper		lower	upper
most	2 Sets	1.663	*1.315*	*2.637*	0.870	*0.744*	*0.973*	0.088	*0.013*	*0.130*	0.192	*0.066*	*0.252*	0.000	*0.001*	*0.103*
	3 Sets	1.426	*1.120*	*3.554*	0.843	*0.537*	*0.953*	0.137	*0.013*	*0.167*	0.278	*0.018*	*0.329*	0.000	*0.003*	*0.135*
	4 Sets	1.930	*1.552*	*3.560*	0.908	*0.598*	*1.010*	0.091	*0.004*	*0.131*	0.241	*0.011*	*0.300*	0.000	*0.002*	*0.095*
	5 Sets	3.453	*2.146*	*4.075*	0.588	*0.502*	*0.907*	0.079	*0.009*	*0.174*	0.000	*0.003*	*0.210*	0.000	*0.001*	*0.090*
many	2 Sets	1.645	*0.829*	*1.849*	0.655	*0.595*	*0.851*	0.016	*0.003*	*0.123*	0.000	*0.005*	*0.317*	0.032	*0.003*	*0.131*
	3 Sets	1.678	*1.211*	*2.139*	1.047	*1.004*	*1.228*	0.057	*0.004*	*0.106*	0.000	*0.002*	*0.235*	0.000	*0.001*	*0.103*
	4 Sets	2.884	*1.862*	*3.096*	1.353	*1.271*	*1.702*	0.000	*0.001*	*0.074*	0.000	*0.003*	*0.325*	0.000	*0.001*	*0.091*
	5 Sets	3.086	*1.970*	*4.015*	1.735	*1.618*	*2.110*	0.056	*0.003*	*0.130*	0.000	*0.002*	*0.265*	0.080	*0.041*	*0.164*

Experiment 4

		w			*theta*			*lambda*			*gamma*			*eta*		
			95% CI			95% CI			95% CI			95% CI			95% CI	
			lower	upper		lower	upper		lower	upper		lower	upper		lower	upper
most	2 Sets	0.464	*0.296*	*0.579*	0.984	*0.965*	*1.036*	0.032	*0.005*	*0.081*	0.000	*0.002*	*0.206*	0.000	*0.002*	*0.191*
	3 H Sets	0.902	*0.593*	*1.236*	1.098	*1.043*	*1.232*	0.025	*0.007*	*0.050*	0.000	*0.005*	*0.415*	0.000	*0.001*	*0.134*
	3 D Sets	0.623	*0.448*	*0.965*	1.067	*1.006*	*1.128*	0.101	*0.005*	*0.161*	0.124	*0.004*	*0.201*	0.000	*0.002*	*0.165*
many	2 Sets	0.580	*0.375*	*0.845*	0.861	*0.828*	*0.937*	0.047	*0.007*	*0.082*	0.000	*0.003*	*0.295*	0.000	*0.001*	*0.162*
	3 H Sets	0.611	*0.203*	*0.853*	1.084	*1.033*	*1.217*	0.046	*0.030*	*0.075*	0.000	*0.004*	*0.399*	0.000	*0.002*	*0.167*
	3 D Sets	1.293	*0.631*	*1.428*	0.813	*0.744*	*1.039*	0.000	*0.002*	*0.111*	0.000	*0.004*	*0.433*	0.054	*0.013*	*0.207*
many (*N* =	2 Sets	0.446	*0.271*	*0.761*	0.936	*0.910*	*1.008*	0.084	0.011	0.130	0.000	*0.002*	*0.254*	0.000	*0.002*	*0.207*
	3 H Sets	0.700	*0.261*	*1.080*	1.101	*1.040*	*1.230*	0.065	0.037	0.104	0.000	*0.004*	*0.408*	0.000	*0.002*	*0.187*
	3 D Sets	0.513	*0.416*	*1.158*	1.068	*0.920*	*1.118*	0.100	0.003	0.144	0.283	*0.005*	*0.330*	0.000	*0.004*	*0.200*

References

Aloni, M., Beaver, D., Clark, B., & Van Rooij, R. (2002). The dynamics of question and focus. *Semantics and Linguistic Theory, (12),* 20-39.

Barner, D. (2012). Bootstrapping numeral meanings and the origin of exactness. Language Learning and Development, 8(2), 177-185.

Barner, D., & Bachrach, A. (2010). Inference and exact numerical representation in early language development. *Cognitive psychology, 60*(1), 40-62.

Barner, D., Chow, K., & Yang, S. J. (2009). Finding one's meaning: A test of the relation between quantifiers and integers in language development. *Cognitive psychology, 58*(2), 195-219.

Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. In *Philosophy, Language, and Artificial Intelligence* (pp. 241-301). Springer: Netherlands.

Blutner, R. (2000). "Some Aspects of Optimality in Natural Language Interpretation," *Journal of Semantics 17*, 189–216.

Boersma, P. & Hayes, B. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry, 32*(1):45–86.

Büring, D. (2003). On D-trees, beans, and B-accents. *Linguistics and philosophy, 26*(5), 511-545.

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

Carlson, G. (1977). Reference to kinds in English. (Doctoral dissertation, University of Massachusetts).

Choi, H., (1996). *Optimizing structure in context: Scrambling and information structure.* (Doctoral dissertation, Stanford University).

Cohen, A., (2001). Relative readings of *many*, *often*, and generics. *Natural Language Semantics, 9*, 41-67.

Cohen, A., & Krifka, M. (2014). Superlative quantifiers and meta-speech acts. *Linguistics and philosophy, 37*(1), 41-90.

Davies, M. (2008-) The Corpus of Contemporary American English (COCA): 150 million words, 1990-present. Available online at http://corpus.byu.edu/coca/.

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1-12. doi:10.3758/s13428-014-0458-y.

Dehaene, S. (1997). *The Number Sense: How the mind creates mathematics*. New York: Oxford University Press.

Dekker, P., & Van Rooy, R. (2000). Bi-directional optimality theory: An application of game theory. *Journal of Semantics, 17*(3), 217-242.

Diesing, M. (1988). Bare plural subjects and the stage/individual contrast. In *Genericity in Natural Language: Proceedings of the 1988 Tübingen Conference* (pp. 107-154).

Feigenson, L., Dehaene, S., & Spelke, E., (2004). Core systems of number. *TRENDS in Cognitive Science, 8(7),* 307-314.

Fara, D. G., (2000). Shifting sands: An interest-relative theory of vagueness. *Philosophical Topics, 28*. 45-81.

Fine, K., (1975). Vagueness, truth, and logic. *Synthese, 30.* 265-300

Freed, B., Dewey, D., Segalowitz, N., & Halter, R., (2004). The Language Contact Survey. *Studies in Second Language Acquisition, 26*, 349-356.

Fründ, I., Haenel, N. V., & Wichmann, F. A. (2011). Inference for psychometric functions in the presence of nonstationary behavior. Journal of vision, 11(6), 16-16.

Ginzburg, J. (1995a). Resolving questions, I. *Linguistics and Philosophy, 18*(5), 459-157.

Ginzburg, J. (1995b). Resolving questions, II. Linguistics and Philosophy, 18(6), 567-609.

Ginzburg, J., (1997). Interrogatives: Questions, Facts, and Dialogue. In. Shalom Lappin (Ed.), *The Handbook of Contemporary Semantic Theory*. Retrieved from http://www.bl.ackwTelhlrefeqreuncees.ctoimo/nsu/biscsriuber/uisid=s9t7il6l/toacnoodep?eidn=go97n8e06.31207498_chunk_g978063120749818&authstatuscode=202

Grice, H. P. (1975). "Logic and conversation" In Cole, P., and Morgan, J.(Eds.). *Syntax & Semantics, 3*.

Hackl, M. (2009). On the grammar and processing of proportional quantifiers: most versus more than half. Natural Language Semantics, 17(1), 63-98.

Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science, 17(7),* 572-576.

Hamblin, C.L., (1973). Questions in Montague English. *Foundations of Language*, *10* (1): 41–53.

Hendriks, P. (2004). Optimization in focus identification. *In Optimality theory and pragmatics* (pp. 42-62). Palgrave Macmillan, London.

Hendriks, P., & De Hoop, H. (2001). Optimality theoretic semantics. *Linguistics and philosophy, 24*(1), 1-32.

Herburger, E. (2000). *What counts: Focus and quantification*. Boston: MIT Press.

Herburger, E., (1997). Focus and weak noun phrases, *Natural Language Semantics, 5,* 53-78.

Horn, L. R., (1984). Towards a new taxonomy for pragmatic inference: Q-based and R-based implicatures', In D. Schiffrin (ed.), *Meaning, Form, and Use in Context*. Washington, D.C.: Georgetown University Press, 11-42.

Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.

Kennedy, C. (2019). Ambiguity and vagueness: An overview. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.) *Semantics-Lexical Structures and Adjectives* (507-535) Boston: De Gruyter.

Kennedy, C. (2011). Ambiguity and vagueness: An Overview. In C. Maienborn, K. van Heusinger, & P. Portner (Eds.) *Semantics*: *An International handbook on Natural Language Meaning* (Vol. 1). Boston: De Gruyter.

Kennedy, C. 2007. Vagueness and grammar: The semantics of relative and absolute gradable predicates. *Linguistics and Philosophy, 30*. 1–45.

Kingdom, F. A.A, & Prins, N. (2016). *Psychophysics: A Practical Introduction*. 2nd ed. London: Elsevier Science.

Kiss, K. É. (2008). Topic and focus: Two structural positions associated with logical functions in the left periphery of the Hungarian sentence. *Acta Linguistica, 55*(3-4), 287-296.

Krifka, M., (1995). Focus and the interpretation of generic sentences. In Carlson, G. & Pelletier, J. (Eds.) *The Generic Book*, (238-264). Chicago: University of Chicago Press.

Krifka, M. (2001). For a structured meaning account of questions and answers. *Audiatur vox sapientia. a festschrift for Arnim von Stechow*, 15, 287-319.

Kuno, S. (1973). *The Structure of the Japanese Language*. Cambridge, MA: MIT Press.

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

Lutken, C. J. (2019). A cross-linguistic analysis of Scope Marking using Optimality Theory. Unpublished manuscript.

Milsark, G., (1977). Toward an explanation of certain peculiarities of the existential construction in English. *Linguistic Analysis, 3*, 1-29.

Pancheva, R., & Tomaszewicz, B. M. (2012). Cross-linguistic Differences in Superlative Movement out of Nominal Phrases. *Proceedings of the 30th West Coast Conference on Formal Linguistics,* ed. by N. Arnett and R. Bennett, 292-302. Somerville, MA: Cascadilla Proceedings.

Parsons, K. P. (1973). Ambiguity and the truth definition. *Nous*, 379-394.

Partee, B. (1988). Many quantifiers. In Proceedings of ESCOL (Vol. 5, pp. 383-402).

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

Prince, A., & Smolensky, P. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar.* Technical Report 2. Center for Cognitive Science, Rutgers University, New Brunswick, NJ

Rett, J. (2014). *The semantics of evaluativity*. Oxford: Oxford University Press.

Roberts, C., (1996/2012). Information structure in discourse: Towards an integrated theory of pragmatics. *Semantics & Pragmatics, 5,* article 6: 1-69.

Romero, M. (2015a). The conservativity of many. In *Proceedings of the 20th Amsterdam Colloquium* (pp. 20-29). Amsterdam: University of Amsterdam.

Romero, M., (2015b). POS, -est and the reverse readings of *many* and *most*. In Proceedings of *NELS* 46.

Rooth, M., (1992). A Theory of Focus Interpretation. *Natural Language Semantics 1*: 75-116.

Schwarzschild, R. (1999). GIVENness, AvoidF and other constraints on the placement of accent. *Natural language semantics, 7*(2), 141-177.

Schütt, H. H., Harmeling, S., Macke, J. H., & Wichmann, F. A. (2016). Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. Vision Research, 122, 105-123.

Smolensky, P., & Legendre, G. (2006). *The harmonic mind: From neural computation to optimality-theoretic grammar (Cognitive architecture)*, Vols. 1 & 2. Boston: MIT Press.

Solt, S. (2009). Notes on the comparison class. In *International workshop on vagueness in communication* (pp. 189-206). Springer, Berlin, Heidelberg.

Solt, S. (2011). Vagueness in quantity: Two case studies from a linguistic perspective. In *Understanding vagueness: Logical, philosophical and linguistic perspectives*, 157-174.

Solt, S. (2015). Q-adjectives and the semantics of quantity. *Journal of Semantics, 32*(2), 221-273.

Stalnaker, R.: 1978, 'Assertion', in P. Cole (ed.), *Syntax and Semantics*, Vol. 9, New York: Academic Press.

Szabolcsi, A. (1986). Comparative superlatives. MITWPL 8, 245–66.

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

Tieu, L., & Shen, Z. (2015). Searching for absolute and relative readings of superlatives: a second experiment. In *BUCLD 39: Proceedings of the 39th Annual Boston University Conference on Language Development, November 7-9*, 2014, Boston University (pp. 424-436).

Tomaszewicz, B. M. (2016). Relative readings of superlatives: Scope or focus?. *Semantics and Linguistic Theory* (Vol. 25, pp. 415-470).

Tomaszewicz, B. (2013). Linguistic and visual cognition: Verifying proportional and superlative most in Bulgarian and Polish. Journal of Logic, *Language and Information, 22*(3), 335-356.

von Fintel, K. (2004). A minimal theory of adverbial quantification. in *Context Dependence in the Analysis of Linguistic Meaning,* Volume 11 of "Current Research in the Semantics/Pragmatics Interface", eds. Barbara Partee & Hans Kamp, Elsevier, pp. 137–175.

von Fintel, K. (1994). *Restrictions on quantifier domains* (Doctoral dissertation, University of Massachusetts).

van Rooy, R. (2003). Questioning to resolve decision problems. *Linguistics and Philosophy, 26*(6), 727-763.

Wagner, M. (2012). Contrastive topics decomposed. *Semantics and Pragmatics*, 5, 8-1.

Westerståhl, D. (1985). Logical constants in quantifier languages. *Linguistics and Philosophy, 8*(4), 387-413.

White, L. (2003). *Second language acquisition and universal grammar*. Cambridge University Press.

White, L., & Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second language research*, *12*(3), 233-265.

Zaroukian, E., (2013). *Quantification and (Un)certainty* (Doctoral dissertation, The Johns Hopkins University).

Vita

Karen Clothier was born in 1989 in Denver, Colorado. After completing her Bachelor of Science at Victoria University of Wellington in New Zealand, she moved to Baltimore and joined the Cognitive Science Department of The Johns Hopkins University. There she completed her Master of Arts in Cognitive Science in 2016. In the fall of 2019, she will begin her role as a Senior Radio Telecommunications Policy and Regulations Analyst at Teltrium Solutions LLC., in Greenbelt, MD, advising and facilitating the telecommunications policy and regulatory processes for NASA and other federal and corporate customers.