What makes *a few* more than *a lot*: a study of context-dependent quantifiers

by

Amanda Pogue

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Arts

in

Psychology

Waterloo, Ontario, Canada, 2013

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

_____

Amanda Pogue

Abstract

"Hey can you help me move? I warn you I have a lot of books, though." When we interpret such sentences we might assume that our friend is implying that there will be some heavy lifting, because she own "a lot of books". If you're opposed to heavy lifting, you probably want to know how many books your friend wants you to help her move.

While it is easy to determine the quantities picked out by numerals, discovering the meaning of quantifiers, such as "a lot", is less clear. For example, in a survey people gave different quantities for "a few" and "a lot" depending on the context they were asked about, choosing to give quantities as high as 76 for "a few friends on Facebook" and as low as 4 for "a lot of houses". I ask what role context plays in these interpretations, and propose two possible hypotheses: the Fixed Quantities Hypothesis, and the Relative Quantities Hypothesis. The Fixed Quantities Hypothesis assumes that these terms pick out an exact range of quantities (e.g., 3-5 for "a few") and that the effect of context is due to people being pragmatically generous in their interpretations (Leech, 1983; Brown & Levinson, 1987; Lasersohn, 1999). The Relative Quantities Hypothesis instead argues that the ranges picked out by each of the terms is relative to the context, and therefore these terms might be similar to relative gradable adjectives (e.g., the meaning of the word "tall" is dependent on the reference set, e.g., what counts as "tall" is different for a mug versus a building; Kennedy, 2007; Syrett, Kennedy, & Lidz, 2010). I attempt to tease apart these two theories by looking at whether people find certain quantities implausible for some contexts leading them to think a speaker is being sloppy (Experiment 1), whether the context effect persists even under light-to-no pragmatic pressure (Experiment 2), and whether we find the effect of context even with minimal knowledge of novel contexts (Experiment 3). The results favour the Relative

Quantities Hypothesis, and I discuss potential future work investigating the role of distributional knowledge on quantifier mappings.

Acknowledgments

Firstly, I would like to thank my advisor Mathieu Le Corre for agreeing to take on this ambitious project with me, providing me with support, and for guiding me through a vast jungle of uncertainty. I would also like to thank my committee members Ori Friedman, and Stephanie Denison, both of whom have been very generous in providing me with advice and feedback. I am also grateful for all of the hard work and effort put into data collection by my two wonderful research assistants, Adel Jalabi and Rachel Oliver. Finally, I appreciate all of the support offered to me by my family, my friends, and my colleagues at the University of Waterloo, especially Amy Yeung, Meghan Dale, Jennifer Collins, Megan Owen, Divam Jain, Christie Haskell, Ashleigh Grace Latimer, Emanuela Yeung, Margaret Echelbarger, Katherine White, Bill Eickmeier, and Rita Cherkewski. Thanks, everyone, for encouraging me to keep on keeping on.

Table of Contents

List of Tables

List of Figures

Introduction

One of the greatest feats accomplished by humans is our ability to communicate using language. Some people might even argue that our linguistic abilities are what separate us from our banana loving, primate relatives (Pinker, 1994; Deacon, 1998). One of the most astounding aspects of language that is both simple, and seemingly unconscious, is our ability to communicate complex information in very few words. For example, imagine yourself at a coffee shop with a friend you haven't seen in a while, and she exclaims, "I'm sorry it has been so long since we caught up, I've travelled a lot this year." Easily you interpret the latter half of this statement to mean that your friend has been away on several trips. However, if you look again at her statement, the literal interpretation cannot tell you explicit information about how many trips she took. So how do we come to the conclusion that she went on more than just one trip?

Words like "a lot" denote a vague sort of quantity. Unlike terms like "one" or "two" which pick out very specific quantities in the world (namely, one, and two, respectively), terms like "a lot" refer to some unspecific quantity. However, in statements such as the one above adults quickly make estimations about what quantity a speaker might mean. Interestingly however, it would appear that the quantity these terms denote does not necessarily remain constant across contexts. For example, consider the following statements:

(1) a. Dave has a lot of friends on Facebook.

b. Dave owns a lot of houses.

(2) a. Dave has five friends on Facebook.

b. Dave owns five houses.

The quantity associated with "a lot" in sentence (1a) might be very different than the quantity associated with "a lot" in (1b), whereas the quantity associated with "five" does not vary between (2a) and (2b). In fact, we might be willing to accept (2b) as a plausible interpretation of (1b), however, it seems highly unlikely that anyone would agree that (2a) is a plausible interpretation of (1a). We conducted a pilot study looking at this intuition. We gave participants a survey where they were asked to estimate the quantities implied by the terms "a few", "a lot", and "average number" in sentences such as (1a) and (1b). They were asked to make estimations for these terms in a variety of different contexts which all reflected common things, such as types of things individuals might own or have (e.g., pets, DVDs, guitars), and things or events in the world (e.g., trees in a park, emails received in a week). Half of the scenarios were expected a priori to have very small averages, and the other half were expected to have relatively higher reported averages. Participants gave a wide range of quantities[1]. We analyzed their responses, and found that there was an effect of context, mainly, that the quantities estimated varied by context[2]. More specifically we found that people gave larger quantities for "a few" in the larger average contexts, than in the smaller average contexts[3], and smaller quantities for "a lot" in the smaller average contexts than in the larger average contexts[4]. From this we can conclude that there is an effect of context on people's interpretation of the terms "a few" and "a lot".

It is important that context play a role in the interpretation of these terms, because while 5000 houses is most certainly "a lot of houses" owned by an individual, it is unlikely that anyone

[1] Outliers were eliminated prior to analysis using a recursive outlier detector on individual test items, and removed any data points that were deemed to be outliers by the PJ Outlier program (Van Selst & Jolicoeur, 1994). Using this measure we deleted 3.6% of the data.
[2] We found an effect of Context ($F(19,95) = 10.172$, $p < .001$), Quantifier ($F(1,5) = 16.987$, $p < .001$) and an interaction of Context and Quantifier ($F(19,95) = 10.334$, $p < .001$).
[3] $t(182) = 7.9977$, $p < .0001$
[4] $t(179) = 8.1881$, $p < .0001$

ever means 5000 houses when they say that someone owns "a lot of houses". Though, technically 5000 houses is a lot of houses owned by a single individual, given that the average individual owns 1 or fewer houses. Similarly, despite the fact that most people have more than 100 friends on Facebook there is no reason why people should not estimate that "a few friends on Facebook" is just 3 friends. However, very few people estimated such quantities on our pilot study. What is particularly interesting about studying the role of context on people's expectations about the actual quantities reflected by these terms is that we get to investigate how people are making these mappings.

What is it about context that is allowing people to create or constrain these context appropriate mappings? We consider two potential hypotheses. The first hypothesis is grounded in pragmatics (or the social conventions used in language), and assumes that the terms have fixed meanings (e.g., the term "a few" always means a quantity between 3-5), and that people are willing to be pragmatically nice in their interpretation of the term when they think that the actual quantities reflected by the terms are implausible for a given context (e.g., it is unlikely that anyone has *only* 3-5 friends on Facebook); we refer to this as the Fixed Quantities Hypothesis. A second hypothesis suggests that the role of context instead is directly linked to the semantic (or logical, truth-evaluable) meaning of the terms. This hypothesis suggests that the terms themselves have no fixed meanings but instead the quantities associated with them are relative to the context being discussed; we refer to this as the Relative Quantities Hypothesis. By teasing apart these two hypotheses this thesis aims to determine the underlying structure of these quantity terms.

The Fixed Quantities Hypothesis

If "a few" and "a lot" each refer to a fixed range of quantities, then why do the quantities associated with each of these terms vary by context? One hypothesis is that people use these terms

sloppily, and listeners are generous when they hear sloppy uses, and accommodate their interpretations. Listeners can use rules of linguistic politeness (Brown & Levinson, 1987) to accommodate for a speaker's sloppiness. Several researchers suggest that people speak with varying degrees of preciseness, and in fact sometimes speak loosely (Sperber & Wilson, 1986; Lasersohn, 1999). This is to be distinguished from truth-conditional vagueness (Kennedy, 2007), which instead focuses on vagueness in meaning, rather than on pragmatic looseness in the usage of a term. Instead pragmatic looseness suggests that in some circumstances exactness is not necessary, so people are free to be looser in their language use. For example, when we talk about time it is not always necessary to communicate the time down to the second. Thus, if I were to message a friend and tell her, "Marc got here at 6:15" and she finds out that in fact Marc arrived at 6:16, it would not be unreasonable for her to protest, but it is unlikely that she would. While it is true that what I said was literally false, for the practical purposes of communication I was "close enough" to the truth for her to let my sloppiness slide. Essentially if the use of a term is close enough to true for a context, a listener should be willing to accept its usage.

Perhaps then, participants in our pilot study were willing to accept that the statements they were reading were sloppily constructed for some contexts, and thus chose to accommodate their assumptions on what we meant to ask about. They might have been pragmatically generous, an aspect of linguistic politeness (Leech, 1983; Brown & Levinson, 1987). In the case of the Fixed Quantities Hypothesis, we are projecting that people are being generous in their interpretations of what a speaker means when they utter a sentence. For example, if we have a strict range set for "a few" to mean approximately 3-5, then a sentence such as (3), might seem not merely sloppy, but rather infelicitous. Rather than correct the speaker, and inform them that it is highly unlikely that Tara only has 3-5 friends on Facebook (participants in our pilot study estimated that on average

people have about 236 friends on Facebook), the listener could be accommodating, and assume that the speaker meant to imply that Tara merely has a small number of friends on Facebook, and interpret the statement as such.

(3) Tara has a few friends on Facebook.

Given this interpretation, what we should expect is that in our prior survey that people were using this line of reasoning for all contexts where 3-5 were implausible quantities. In Experiment 1, we look at whether people really find these quantities implausible for contexts that typically have large overall averages. Presumably, if the Fixed Quantities Hypothesis is correct we should expect people to prefer to give the same quantities for "a few" in all contexts, and when they do not, we should predict that people will rate quantities between 3 to 5 as implausible for those contexts.

The Relative Quantities Hypothesis

Unlike the Fixed Quantities Hypothesis, the Relative Quantities Hypothesis does not assume that the effect of context on the quantities associated with "a few" and "a lot" is due to a social convention. Instead, on this hypothesis, the context-dependency is built into the very meaning of the terms – they *must* be interpreted within a particular context to refer to a quantity. In other words, in the absence of any context, the quantities associated with the terms are only partially defined, or not defined at all. Moreover, the quantities associated with the terms are sensitive to the distributions of quantities that are typical of the contexts where they are used. They are small when the terms are used in contexts that are typically associated with small quantities, and they are larger in context that are typically associated with large quantities.

In this sense, the Relative Quantities Hypothesis treats "a few" and "a lot" much like relative gradable adjectives – words such as "tall", "long", and "small". Gradable adjectives are said to be

terms that denote some sort of measurable property of an individual or substance, such as size or volume. Specifically relative gradable adjectives are terms that denote measurable properties that require the listener to take into consideration a reference set when determining what it picks out in the world, and whether or not it was used felicitously (Kennedy, 2007; Syrett, Kennedy, & Lidz, 2010). Under this definition a term like "tall" denotes no specific height, and can only do so when making reference to a particular set of things.

Consider the following sentences:

(4) Meet me by the tall tree in the park.

(5) Corey is really tall (for a man).

There are two circumstances when "tall" can be used meaningfully: first, when in reference to a particular set of things, such as in (4), and second, when in reference to an object with respect to its kind, as in (5). In (4) the listener can use "tall" to pick out a specific tree in the park, by searching for the tallest tree out of all of the trees in the park. In this case the word "tall" requires the listener to compare individuals to a specific set, namely the set of trees in the park. Otherwise in (5) the sentence is felicitous only if it is the case that Corey is in fact "tall" for a man. Thus, in (5) the listener has to compare Corey's height to the average height of men in general, and consequently the reference set in this scenario is the set of all men.

Likewise the Relative Quantities Hypothesis posits that terms like "a few" and "a lot" have a similar structure; meaning that they also do not pick out a specific range, and instead get their range from the context or reference set. It is possible that for terms like "a few" and "a lot" people use information about the distribution and / or average information for that context when making their decision about the intended quantity.

Based on this hypothesis, we ought to predict that participants will be willing to give a variety of different ranges for each of the terms depending on the context. Specifically, this hypothesis predicts the findings we found in our pilot study showing an effect of context on quantities given for "a few" and "a lot". It would also predict that these findings are not due to pragmatic adjustments as suggested by the Fixed Quantities Hypothesis. Thus, the Relative Quantities Hypothesis makes a different prediction for Experiment 1. Instead it would predict that people are not assuming that the speaker is being sloppy in their use of "a few" for larger average contexts (such as in, "a few friends on Facebook"), and thus it predicts that they would not find quantities within the acceptable range of "a few" as defined by the Fixed Quantities Hypothesis to be implausible interpretations of the term "a few" for those contexts. Further, the Relative Quantities Hypothesis would suggest that the effect of context should appear even when people are not put under pragmatic pressure to be polite, as will be investigated in Experiment 2. Finally, it predicts that if people are sensitive to distributional information, then we should still expect to see variability based on context information when people are given minimal exposure to novel contexts. We explore this possibility in Experiment 3.

Overall this thesis will look at each of the hypotheses across three different studies, and determine which best fits how adults interpret these terms. For the purpose of this thesis we discuss two terms: "a few" and "a lot". In the discussion we will consider the implications of these findings for the purposes of using language, and for acquisition.

Experiment 1

Given that we have seen an effect of context on the quantities associated with the terms "a few" and "a lot" in our pilot study, we must seek an explanation that can account for the variances in the responses. The Fixed Quantities Hypothesis suggests that the terms represent quantities with strict boundaries, but the reason why we find such varied effects of context in our pilot study is that people are willing to accept or give larger quantities out of pragmatic politeness.

This explanation assumes that participants in our pilot study find the uses of the terms "a few" and "a lot" infelicitous for certain contexts, as they do not believe it plausible that the quantities associated with each term are likely for those contexts. In order to answer the questions posed to them, the participant must then be generous in their interpretation of the terms and assume that the speaker meant something more like "a small (plausible) quantity" for "a few" and "a large (plausible) quantity" for "a lot". If this explanation were correct, then we should predict, for example that people would rate sentences that use quantities within the range of "a few" to be implausible for contexts that normally involve large quantities (e.g, the typical number of friends people have on Facebook), indicating that people are in fact using this politeness correction strategy. In Experiment 1 we provided participants with statements such as in the pilot study, but instead of using the terms "a few", "a lot", and "average" each statement included a numeral between 3 to 5, and we asked participants to rate the plausibility of each statement. If the Fixed Quantities Hypothesis is correct, then we should expect variability in the plausibility ratings, and should particularly expect low plausibility ratings for all contexts that normally involve large quantities.

Method

Participants. Forty-four fluent English-speaking adults (M = 20.4 years old, 25 females) were recruited from, and participated at the University of Waterloo. Participants were recruited online using the REG-SONA system, received partial course credit for participating, and were naïve as to the purpose of the experiment.

Materials and Procedure. Participants were given a questionnaire. The questionnaire contained two surveys: an estimation survey, and a plausibility survey. The estimation survey was identical to the design of our pilot study, except that only the quantifier, "a few" was included. Participants were given a series of questions with seven new contexts. These contexts included: silver cars in a parking lot, trips taken to the mall / gym / library in a semester, upper year students in a first year class, movies watched in a semester, and instances of eating pizza in a year. We predicted that these contexts would represent a set of test items where it is both plausible for the fixed range of "a few" (3-5), as well as larger numbers to be used to describe these contexts in the real world (e.g., it is plausible that there are only 5 silver cars parked in the student parking lot today, or there could be 20 silver cars in the lot). For each item participants were asked to estimate the quantity implied by "a few". For example, for silver cars they were asked, "There are a few silver cars in the UW parking lot today. How many silver cars do you think are in the UW parking lot?" Two item orders were created, and participants were randomly assigned to each order.

After completing the estimation survey participants were instructed to complete a second survey; the plausibility survey. The plausibility survey included the 7 contexts from the estimation survey and 9 of the contexts from the pilot study (apples picked on a trip, books owned by an individual, trees in a park, DVDs owned by an individual, emails received in a week, friends on Facebook, cars owned by an individual, photos tagged online, pennies in a jar, and shirts owned

by an individual). Each of the contexts were presented in sentences as in the estimation survey, but instead of using the term "a few", the new sentences incorporated a number between 3 to 5, which we deemed to be the approximate range for "a few". For example, they were given sentences such as, "Megan is tagged in four photos online." Participants were given the following instructions:

> For the following section please rate the plausibility of the following sentences on a scale of 1 to 5, where 1 is highly implausible, and 5 is very plausible.

The sentences were presented in a list. For each sentence participants were given a scale between 1 to 5, and they indicated their response by circling the number they thought was appropriate on the scale. Two item orders were created; the old and new items were interspersed within each order. Participants always received the plausibility survey after receiving the estimation survey. Each survey was printed on a separate sheet of paper.

Results and Discussion

We conducted an outlier analysis to remove any responses to the estimation survey that were greater than 2.5 standard deviations above or below the mean for each item[5]. Table 1 reports the average quantities given for the estimation survey, and the pilot study for "a few," as well as the average plausibility ratings for all the test items rated on the plausibility survey. The items are listed in decreasing order of plausibility. Table 1 shows that subjects associated quantities larger than 5 with several of the items that were rated as having at least medium plausibility (3 out of 5 or higher). Only 3 of the 16 items tested on the plausibility survey did not achieve plausibility ratings significantly above 3 (neither plausible, nor implausible; *p*s > .01); number of trees in a

[5] Using this measure we deleted 6.82% of the data

park, number of friends on Facebook, and number of silver cars found in a parking lot. Just one context, number of friends on Facebook, fell significantly below 3 ($t(42)= -2.118$, $p < .05$). Despite the fact that participants gave quantities outside of our predicted range for "a few", they did not overall reject quantities within the range as being implausible. This provides some evidence against the Fixed Quantities Hypothesis.

There are three flaws with the design of this particular experiment. First, unfortunately since we only collected estimation data for "a few" in this study, and only for the new items we cannot run a correlation between "average" estimations and "a few" estimations with plausibility partialled out. If we found a significant correlation, we would have much stronger evidence against the Fixed Quantities Hypothesis. Second, this experiment assumed that the potential fixed range of "a few" was between 3 to 5; it is possible that this is not an accurate picture of the actual range, and thus our data may not reflect the true fixed range. Finally, the design of this experiment could have benefitted from counterbalancing the order of the surveys, as the current order of receiving the estimation survey first may have influenced participants' plausibility judgments on the second survey.

Table 1. *Mean responses given for "a few X" for items from the pilot study and Experiment 1, and the mean plausibility rating for each corresponding trial in Experiment 1.*

Test Item	Plausibility of 3-5	A Few
E-mails in a week	4.42**	8.39
Trips to the mall in a semester	4.42**	4.02
Trips to the gym in a year	4.28**	8.18
Trips to the library in a semester	4.28**	5.73
Upper year students in a first year class	4.19**	10.61
Movies watched in a semester	4.19**	6
Apples picked in an outing	3.94**	8.28
Books	3.89**	10.16
Photos tagged online	3.78**	27.16
Pennies in a jar	3.78**	25
Shirts	3.75**	10.89
Pizzas eaten in a years	3.75**	8.48
DVDs	3.67*	8.28
Silver cars in the parking lot	3.19	8.1
Trees in a park	3.08	18.5
Friends on Facebook	2.5	76.26

Note: significance is reported for comparison to a plausibility score of 3 (*$p < .01$, ** $p < .001$)

Experiment 2

In Experiment 1 we attempted to investigate the role of context in people's interpretations of the quantity term "a few". We investigated the possibility that people have a fixed range of quantities picked out for the term "a few" and that they are willing to use pragmatic linguistic strategies to accept terms outside of that fixed range. The Fixed Quantities Hypothesis predicts that there is an acceptable range of quantities that people are willing to associate with the term "a few", and that there might be some contexts where quantities within this range just seem implausible in the real world. The findings of Experiment 1, suggested that these predictions were no supported. However, the findings are based on the assumption that the fixed range for "a few" is actually 3 to 5. The design of the current study makes no predictions about the actual ranges of "a few" and "a lot".

In the pilot study, and in Experiment 1 participants were required to estimate the quantities that should be associated with "a few" and "a lot" for different contexts. If the Fixed Quantities Hypothesis is correct, then people likely felt some pragmatic pressure to give an answer in those tasks. Specifically, since they had to give an answer, if they thought that quantities within the fixed ranges of either term was implausible for a given context, they would still feel required to give an answer. Consequently, when faced with the requirement to pick a response, participants might defer to picking the quantity that they find to be the best possible interpretation of what they are being asked. In Experiment 2, we attempt to reduce this pragmatic pressure. Presumably, under less pressure, participants will be free to reject quantities outside of the fixed ranges for the terms. Instead of asking participants to estimate what quantities "a few" and "a lot" represent in a variety of contexts, we present participants with sentences where "a few" and "a lot" are used to represent different quantities and ask them to rate whether or not the sentences are acceptable. We predict

that this new design reduces the pragmatic pressure to accept quantities outside of the fixed ranges of the terms, as the design itself requires participants to essentially reject poor uses or accept appropriate used of the terms.

Thus, the Fixed Quantities Hypothesis predicted that participants will uniformly accept and reject the same quantities across all contexts for each of the terms (e.g., for "a few" participants ought to accept certain quantities, such as 3-5, for all contexts, and reject all other quantities). Alternatively, the Relative Quantities Hypothesis makes the prediction that the quantities people are willing to accept for each of the terms should vary by context, and thus, the quantities that people are willing to accept as good uses of the terms should vary between contexts.

Method

Participants. Seventy-eight native English-speaking adults (M = 33.8 years old, 44 females) were recruited through Amazon's Mechanical Turk. An additional 6 participants were excluded due to not being native speakers of English (n=1), or for getting 3 or more control questions incorrect (n=5). Participants were paid $2.00 to participate, and were naïve as to the purpose of the experiment.

Materials and Procedure. Participants read through an information page and consented to participate in the study. Each trial contained two sentences: an Original Sentence, and a Test Sentence. The Original Sentence presented a statement including a numeral, for example, "Jason has 50 friends on Facebook." The Test Sentence was identical to the Original Sentence, except the numeral was replaced with a different quantity term, for example, "Jason has *a lot* of friends on Facebook." Participants were asked to compare the sentences.

They were given the following instructions:

> The next few questions will be practice trials. You will be asked to read a sentence, and then on the next page you will be given a second sentence. The second sentence is a description of the first (original) sentence. You will then be asked to rate how acceptable the second sentence is based on the first sentence (e.g., compare the two sentences). Please try to answer this as quickly and accurately as possible. Let's try.

Participants were asked to respond using a 5-point Likert scale where 1 was "Unacceptable" and 5 was "Acceptable". Each participant received two practice trials to get used to the method. The first practice trial was blatantly false (original sentence: Ashleigh has a pet *cat*. Test sentence: Ashleigh has a pet *dog*), and the second practice trial was obviously true (a direct repetition of the original sentence). Participants were then instructed that the task would begin.

One hundred and fifty sentence pairs were created for this task. We used two kinds of quantity terms as test items. These included four quantifiers (a few, a lot, a couple, dozens), and two terms using gradable adjectives (small number, larger number). We included an additional 25 control Test Sentences which replaced the numeral from the Original Sentence with a range of quantities that either contained (e.g., "Jason has *30-60* friends on Facebook"), or did not contain (e.g., "Jason has *20-30* friends on Facebook") the numeral from the Original Sentence.

Each quantity term was compared to five different numbers. The quantities used in the Original Sentences to compare with each quantity term can be seen in Table 2.

Table 2. *Quantities used in Original Sentences to compare with each term.*

Term	Quantities				
A few	2	3	5	10	20
A lot	3	5	10	50	750
A couple	2	3	4	5	10
Dozens	10	20	40	80	150
Small Number	2	3	5	10	20
Large Number	3	5	10	50	750

We asked participants to make these comparisons for five different contexts. The five contexts used in this experiment were from the original pilot study. These included two contexts that have relatively small averages (smartphones, guitars), one with medium average (shirts), and two larger averages (pennies in a jar, friends on Facebook). We used a 5 (Numbers) x 5 (Contexts) design, thus each participant made 25 comparisons per term.

We constructed four Conditions, each with two item presentation Orders. Each Condition contained 100 sentence pairs, which were comprised of all of the sentences pairs from 4 out of 6 of the quantity terms, and all of the control questions. Participants were randomly assigned to a Condition and an Order.

Results and Discussion

We conducted 6 mixed model ANOVAs looking at each of the quantity terms separately. In each ANOVA we checked for effects of Condition (each condition had 2 order, which resulted in 16 Conditions (8 Conditions x 2 Orders)), Gender, Number (quantifier-quantity association), and Context (the context scenario used). We will discuss each quantity term separately.

We are particularly interested in effects of Number and Context, and also the interaction between them. These are really the only theoretically interesting effects in this analysis. The effects of Condition and Gender on their own, or in interactions do not provide evidence for or against the view that the quantities associated with quantifiers depend on context. While effects that involve Condition or Gender on their own, or in interactions with each other will be reported, we will only address them if they appear to be meaningful or relevant.

In terms of theoretically interesting main effects, we will report and analyze any main effects of Number, Context, and their interaction. An effect of Number would suggest that the pattern of results is at least in part driven by individual responses to the different quantities used in the Original Sentence. An effect of Number on its own would suggest that participants are uniform in responding across Contexts. An effect of Context would suggest that the pattern of results is at least partially driven by individual responses that differ across the different Contexts. A significant interaction of Number by Context might hint at the possibility that participants' responses vary depending on the combination of both Number and Context. However, planned comparisons are necessary to determine exactly what is driving the interaction.

A Few

We found a significant main effect of Number ($F(2.3,52.904)=33.457$, $p < .001$)[6], Context ($F(2.662,61.216)=23.787$, $p < .001$), and a two way interaction of Number by Context ($F(16, 368)=3.469$, $p < .001$). There was also a significant interaction of Context x Gender x Condition ($F(18.631,61.216)=1.846$, $p <.05$). No other effects or interactions were significant ($ps > .1$). These findings suggest that participants gave different ratings depending on the numbers given in the

[6] The within-subjects factors of Number and Context failed the test of sphericity, and therefore all results are reported with Greenhouse-Geisser corrections.

Original Sentence, and also based on the Context at hand. It is most interesting to investigate the Number x Context interaction.

Given that our two hypotheses made different assumptions about the role of number and context in this experiment we decided *a priori* that we would need to investigate any potential interactions of Number and Context by determining if the range of acceptability was constant or variable across contexts. We thus decided to determine the highest possible Number that participants were willing to accept for "a few", and found that the number differed depending on Context. For example, 10 friends on Facebook is considered to be significantly above 3 (or neither unacceptable nor unacceptable; $t(38)=3.777$, $p < .002$), but 10 smartphones owned by an individual was considered to be significantly below 3 ($t(38)=-3.571$, $p < .001$). The results of these comparisons can be seen in Table 3. The main finding of this analysis suggests that while 5 is considered to be "a few" for all contexts, it is only acceptable to use "a few" for 10 items in 2 out of 5 contexts. While we can see that the acceptability ratings decrease with an increase in Number, we can also see that it does so unevenly across contexts. We consider the evidence of a Number by Context interaction to be some support for the Relative Quantities Hypothesis for the term "a few".

Table 3. *Mean acceptability ratings given for each Context by Number for "a few" in Experiment 1.*

	2	3	5	10	20
Smartphones	3.31	4.38**	3.85**	2.28**	1.87**
Guitars	3.59*	4.54**	4.08**	2.79	2.05**
Shirts	3.62**	4.44**	4.31**	3.41	2.13**
Pennies in a jar	3.54*	4.72**	4.41**	3.62*	2.56
Friends on Facebook	3.67**	4.44**	4.56**	3.74**	2.77

Note: significance is reported for comparison to a acceptability score of 3 (*$p < .05$, ** $p < .005$)

A Lot

The ANOVA reflected a main effect of Number ($F(2.883,72.079)=432.768$, $p < .001$)[7], Context ($F(4,100)=98.055$, $p < .001$), Order ($F(1,25)=5.615$, $p < .05$) and an interaction of Number by Context ($F(7.136,178.397)=19.498$, $p < .001$). There was also a Context x Gender interaction ($F(4,100)=6.849$, $p < .001$), a Context x Condition interaction ($F(28,100)=2.428$, $p < .01$), a Number x Context x Condition interaction ($F(49.951,178.397)=1.667$, $p < .01$).

We ran the same planned comparisons as with "a few" to investigate the Number by Context interaction, and looked at what the smallest quantity participants were willing to accept for "a lot" for each Context. We did this by comparing the mean acceptability ratings to the middle rating, 3 (neither acceptable nor unacceptable). The results of these comparisons can be seen in Table 4. What we find is that 5 is an acceptable quantity for "a lot" for Smartphones, but no other Contexts, and 10 is an acceptable quantity for "a lot" for both Smartphones and Guitars, but

[7] The within-subjects factors of Number and the interaction between Number and Context failed the test of sphericity, and therefore all results are reported with Greenhouse-Geisser corrections.

nothing else. Also, we find that as the Number increased so does the acceptability rating. However, the acceptability rating starts increasing at smaller quantities for the small average Contexts (smartphones, and guitars) than the other Contexts. Again, these results favour the Relative Quantities Hypothesis, given that we find differences in the smallest quantities considered to be "a lot" across contexts.

Table 4. *Mean acceptability ratings given for each Context by Number for "a lot" in Experiment 1.*

	3	5	10	50	750
Smartphones	2.38**	3.72**	4.23**	4.79**	4.9**
Guitars	1.87**	2.9	3.85**	4.87**	4.97**
Shirts	1.87**	1.41**	2.36**	4.59**	4.82**
Pennies in a jar	1.5**	1.26**	1.82**	3.95**	4.95**
Friends on Facebook	1.08**	1.28**	1.82**	3.69**	4.95**

Note: significance is reported for comparison to a acceptability score of 3 (*$p < .05$, ** $p < .005$)

A Couple and Dozens

Unlike "a few" and "a lot" we predicted that the terms "a couple" and "dozens" would not be context sensitive, and therefore would be clear contrasts with "a few" and "a lot". Specifically, we thought that they might be good evidence for the kind of quantity words that would exemplify the Fixed Quantities Hypothesis. Based on this prediction we expected to find no effect of Context for these quantity terms.

We found this to be true for "dozens" but not for "a couple". Specifically we found no effect of Context ($F(4,92)=.996$, $p > .1$) for "dozens" nor an interaction of Context by Number ($F(7.472,171.856)=1.220$, $p > .1$)[8], but we did find an effect of Number ($F(2.269,52.196)=95.542$, $p <.001$). There was also a significant Number x Gender interaction ($F(2.269,52.196)=3.771$, $p < .05$), Number x Condition x Gender interaction ($F(13.616,52.196)=2.663$, $p < .01$), Condition effect ($F(7,23)=2.987$, $p < .05$). When the data is split by Condition we still find that Number is significant within each condition (ps $< .001$) except in one Condition ($F(4,12)=1.485$, $p > .1$), in the same condition we find a Number x Gender interaction ($F(4,12)=11.005$, $p < .01$), but no where else. We also find that both Context and the interaction of Context and Number remain not significant (ps $> .1$).

We found an effect of both Context ($F(2.789,64.141)=10.690$, $p < .001$)[9] and Number ($F(2.117,48.702)=85.404$, $p < .001$), as well as an interaction between Context and Number ($F(7.529,173.166)=2.929$, $p < .001$) for "a couple".

First we looked at the main effect of Number for "dozens". As in "a few" and "a lot" we compared all of the mean ratings to the middle score of 3 (neither acceptable nor unacceptable). The results are posted in Table 5. As expected, for a main effect of Number, participants are unwilling to accept quantities including, and under 20 as being "dozens" uniformly across all Contexts. Similarly they are uniformly willing to accept 40 and above as being "dozens" across all Contexts.

[8] The within-subjects factors of Number and the interaction between Number and Context failed the test of sphericity for "dozens", and therefore all results are reported with Greenhouse-Geisser corrections.
[9] The within-subjects factors of Number, Context, and the interaction between Number and Context failed the test of sphericity for "a couple", and therefore all results are reported with Greenhouse-Geisser corrections.

Table 5. *Mean acceptability ratings given for each Context by Number for "dozens" in Experiment 2.*

	10	20	40	80	150
Smartphones	1.35**	2.48*	4.48**	4.33**	4.15**
Guitars	1.36**	2.4*	4.38**	4.18**	4.08**
Shirts	1.5**	2.23**	4.25**	4.5**	4.13**
Pennies in a jar	1.3**	2.08**	4.21**	4.2**	4.03**
Friends on Facebook	1.45**	2.23**	4.15**	4.53**	3.98**

Note: significance is reported for comparison to a acceptability score of 3 (*$p < .05$, ** $p < .005$)

Next we looked at the effects and interaction found for "a couple", by conducting some post-hoc analyses by comparing the mean responses for each Context with each Number to the middle response (neither acceptable nor unacceptable). The results can be seen in Table 6. We found that the number 2 was overwhelming accepted as "a couple" for all contexts ($ps < .001$). The responses for the number 3 did not differ from the neutral response, with exception of the "pennies in a jar" context ($ps > .005$). From the quantity 4 onwards all quantities and contexts failed to differ from the middle response (or were significantly below the middle response). Based on these findings it seems like the potential effect of Context seen in this analysis might be driven by differences between contexts based on how unacceptable they are for the larger quantities. Based on this finding it seems plausible to conclude that "a couple" has a sort of fixed definition, in the sense that it appears to have a context independent upper bound of 3, however, it also seems a bit flexible, and that quantities greater than 3 are considered less acceptable for smaller average contexts.

Table 6. *Mean acceptability ratings given for each Context by Number for "a couple" in Experiment 2.*

	2	3	4	5	10
Smartphones	4.82*	3.24	2.55	2.32*	1.55*
Guitars	4.92*	3.39	2.68	2.11*	1.58*
Shirts	4.92*	3.42	2.92	2.63	2.11*
Pennies in a jar	4.74*	3.68*	3.29	3.0	2.26*
Friends on Facebook	4.84*	3.5	3.13	3.0	2.5

Note: significance is reported for comparison to a acceptability score of 3 (* $p < .005$)

Small Number and Large Number

Given the fact that the adjectives "small" and "large" are considered to be relative gradable adjectives (see: Syrett, Lidz, & Kennedy, 2010), and that we found evidence for the Relative Quantifier Hypothesis for "a few" and "a lot", we can predict similar results for "small number" and "large number". We find an effect of Context ($F(2.095,46.098)=66.624$, $p < .001$), and Number ($F(2.402,52.850)=138.484$, $p < .001$), as well as an interaction between Context and Number ($F(7.016,154.341)=14.914$, $p < .001$)[10] for "small number". In addition, there was a Context x Gender interaction ($F(2.095, 46.098)=5.373$, $p < .01$).

For "large number" we found an effect of Context ($F(2.015,48.352)=82.206$, $p < .001$), Number ($F(2.669,64.048)=509.890$, $p < .001$), and a Context x Number interaction ($F(8.322,199.73)=17.825$, $p < .001$). There also was a Number x Condition interaction ($F(18.681,$

[10] The within-subjects factors of Number, Context, and the interaction between Number and Context failed the test of sphericity for both "small number" and "large number", and therefore all results are reported with Greenhouse-Geisser corrections.

64.048)=2.463, $p < .01$), Context x Gender interaction ($F(2.015, 48.352)=5.899$, $p < .01$), and a Number x Context x Condition interaction ($F(58.255, 199.73)=1.408$, $p < .05$). There was also a main effect of Gender ($F(1, 24)=4.745$, $p < .05$), and Condition ($F(7,24)=2.850$, $p < .05$).

When split by Gender we find that there is still an effect of Context, Number, and interaction of Number by Context for both males and females (all ps < .001), the only difference is that there appears to be a Condition effect for females ($F(7,15)=4.263$, $p <.01$) resulting in a Number x Condition interaction (17.784,38.108)=2.468, $p < .05$) and a Context x Condition interaction ($F(16.550,35.463)=2.220$, $p < .05$). When split by Condition we find that there are two Conditions out of eight, where there is no effect of Context (ps > .1) nor an interaction of Number x Context (ps > .05). There was also a Gender effect ($F(1,2)=90.750$, $p < .05$) for one of the two Conditions. The two exceptional Conditions shared test questions for both "large number" and "a couple", but one tested for "a few" and the other tested for "a lot". Furthermore, for each of the exceptional Conditions there was an alternate order where the effects of Context and the interaction of Context x Number were found.

As with "a few" and "a lot" we expected to find an interaction, and thus ran the same planned comparisons looking at the largest quantities that were accepted as being a "small number" for each Context, and what we found was that the largest acceptable quantity varied by Context. The results are in Table 7. Thus providing evidence that the term "small number" is considerably relative.

Table 7. *Mean acceptability ratings given for each Context by Number for "small number" in Experiment 2.*

	2	3	5	10	20
Smartphones	4.59**	4.31**	3.28	2.03**	1.46**
Guitars	4.51**	4.59**	3.95**	2.56*	1.51**
Shirts	4.77**	4.79**	4.49**	3.56*	2.10**
Pennies in a jar	4.64**	4.72**	4.87**	4.31**	3.51*
Friends on Facebook	4.72**	4.77**	4.79**	4.33**	3.38

Note: significance is reported for comparison to a acceptability score of 3 (* $p < .05$, ** $p < .005$)

We ran the same analysis for "large number" looking for the smallest quantity considered acceptable, and also found that the results varied by Context supporting the hypothesis that "large number" is also relative quantity term (results reported in Table 8).

Table 8. *Mean acceptability ratings given for each Context by Number for "large number" in Experiment 2.*

	3	5	10	50	750
Smartphones	1.92**	3.23	4.13**	4.74**	4.95**
Guitars	1.74**	2.92	3.54*	4.74**	5.0**
Shirts	1.21**	1.54**	2.46*	4.74**	5.0**
Pennies in a jar	1.10**	1.08**	1.59**	3.69**	4.92**
Friends on Facebook	1.13**	1.13**	1.85**	3.82**	5.0**

Note: significance is reported for comparison to a acceptability score of 3 (* $p < .05$, ** $p < .005$)

The evidence from Experiment 2 concludes that even when the pragmatic pressure to accommodate a possible interpretation is reduced people are still willing to treat the terms “a few” and “a lot” similar to the gradable adjective based quantity terms “small number” and “large number”, rather than to more fixed terms like “a couple” and “dozens”. These findings support the Relative Quantities Hypothesis on the effect of context on the interpretation of “a few” and “a lot”.

Experiment 3

Experiment 2 provided fairly compelling evidence for the Relative Quantities Hypothesis for the terms "a few" and "a lot". Based on the findings it would appear that people are willing to accept varying ranges of quantities for each term, for different contexts. However, a more convincing story could be told if we found these effects for novel contexts. In Experiment 3, we provide participants with several novel contexts, and minimal distributional information about the contexts, and then ask them to estimate what counts as "a few" and "a lot" for each of the contexts. If the Relative Quantities Hypothesis is correct, then we should predict that when given minimal information about distributional information for novel contexts, we should still find an effect of context on the quantities people associate with "a few" and "a lot". Again, the two hypotheses make opposite predictions; the Fixed Quantities Hypothesis instead predicts that there will be no effect of contextual information on the quantities associated with the terms.

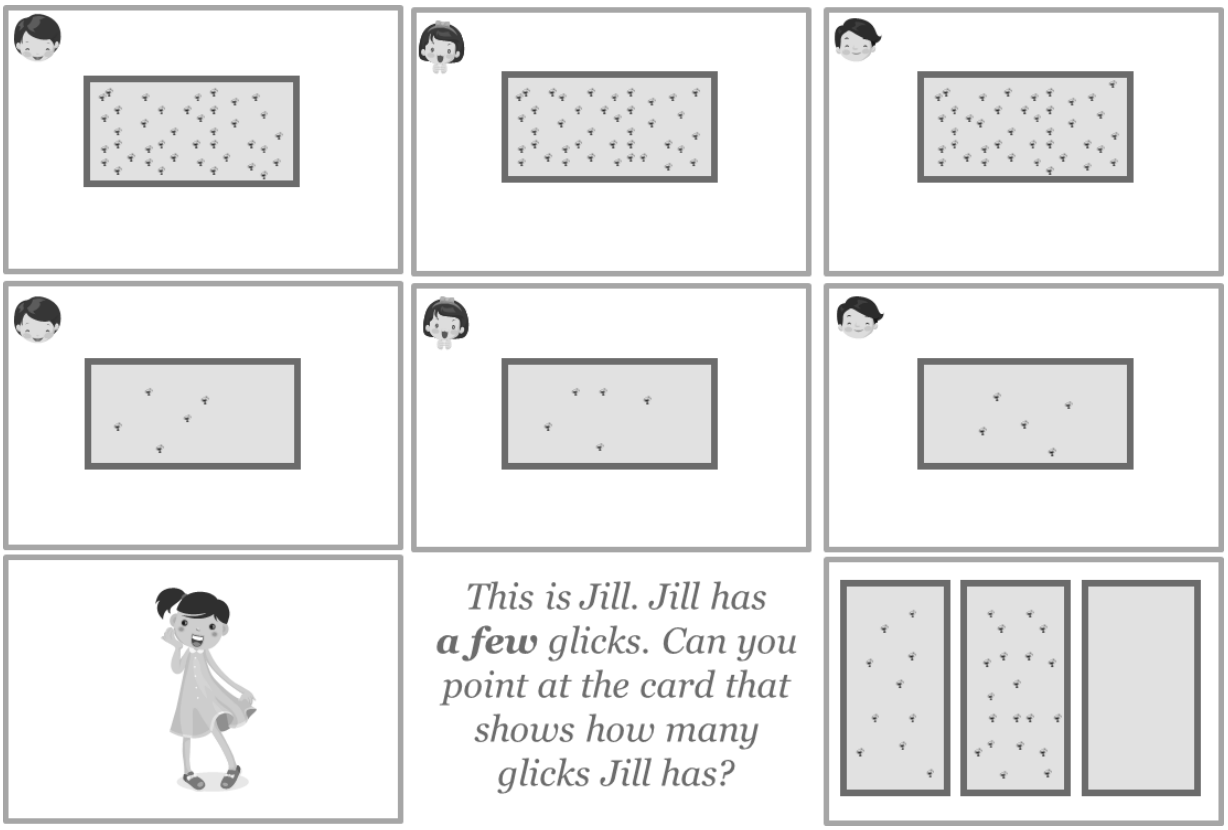
Method

Participants. Thirty-two fluent English-speaking adults (M = 21.1 years old, 17 females) were recruited from, and participated at the University of Waterloo. Participants were recruited online using the REG-SONA system, received partial course credit for participating, and were naïve as to the purpose of the experiment.

Materials and Procedures. Participants were seated at a computer table beside an experimenter. They were briefed on the structure of the task. The participant was instructed that they would be told some short stories and that they would have to respond to questions about the stories. They were then introduced to the answering method; they were shown 3 "cards" on a computer screen and were told that after each question two of the cards would have answers on them, and that these cards were flipped face-up, and the third card would remain blank, or face-

down (for example, see the third panel in Row 3 of Figure 1). If they thought that neither of the face-up cards were correct, they were instructed that the face-down card must have the correct answer. Participants were given three practice trials on the design of the answering method. On these trials participants were asked to locate the card that showed specific items (e.g., a toy car, teddy bear, 2 monkeys), and received feedback on their choices.

Figure 1. *Example of distribution trials for the 40 (Row 1) and 5 (Row 2) conditions, as well as the critical trials (Row 3) in Experiment 3.*



Participants then saw a second set of training trials that introduced participants to the structure of the stories. They were introduced to a novel object, and were told that this particular object was a thing that people in Southern Mexico like to collect. It was predicted that both the

object and the context should be unfamiliar to the participants. Participants were then told a story about three individuals (distribution trials), and were pictorially shown the quantities of these objects that they each owned (2 or 3). Each individual in the story owned the same number of objects. Participants were then asked to estimate how many a fourth person had, and were given the choice between two cards, one with the same number of items as the individuals in the story, and another amount that was either more or less than they had previously seen. They received two training trials of this type, and no corrections were made on these trials.

The test trials included 3 distribution trials, and 1 critical trial. In the distribution trials participants were told stories about a novel object that people in Southern Mexico like to collect, about three individuals, and how many objects each of those individuals owned (see: Rows 1-2 of Figure 1). In the critical trial, participants were introduced to a fourth character, and were told that this character "has a few glicks" or "has a lot of glicks". They were then given the three card answer task, and were asked to pick the card that showed how many glicks they thought the fourth character owned. Row 3 of Figure 1 depicts a typical critical trial.

Participants received 4 test trials, of which half were about characters who on average owned a small quantity of the novel stimulus, and the other half included characters that owned on average a large quantity of the stimulus. Each story contained a different novel object (glicks, pimwits, zubas, mows). Participants were randomly assigned to one of two Quantity Conditions: half of the participants heard stories where the small average quantity was 5 and the large average quantity was 40 (5/40 Context Condition), and the other half heard stories where the characters had 3 (small average quantity) or 30 (larger average quantity; 3-30 Context Condition) of the items. Participants had the option to pick between three cards: 10, 20, and blank (5-40 Context Condition), or 5, 10, and blank (3/30 Context Condition) for each of the stories. If the participant

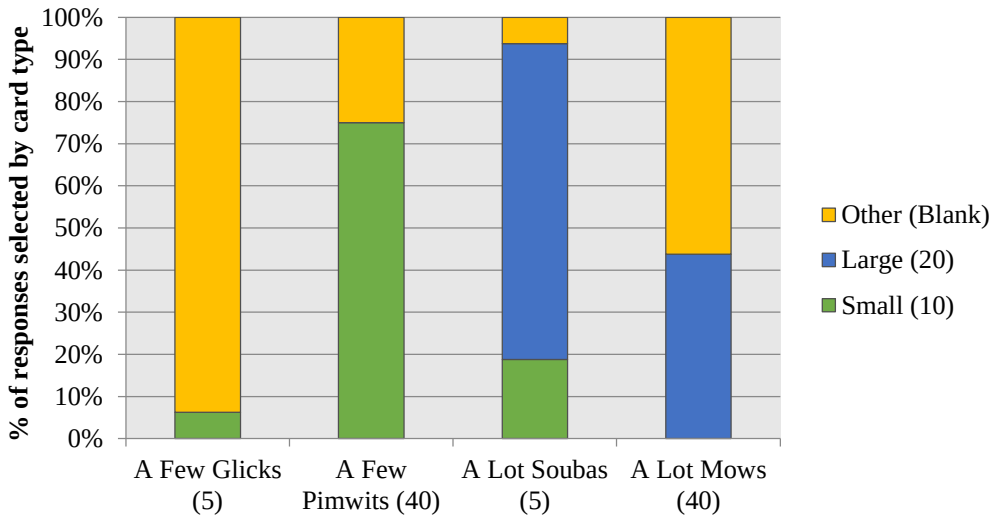
chose the blank card they were elicited to estimate the quantity that they thought might be on the reverse side of the card. Four orders were generated for each of the Quantity Conditions. Each order attempted to counterbalance the arrangement of the selection cards, and the order of the quantity terms. Participants were randomly assigned to a condition.

Results and Discussion

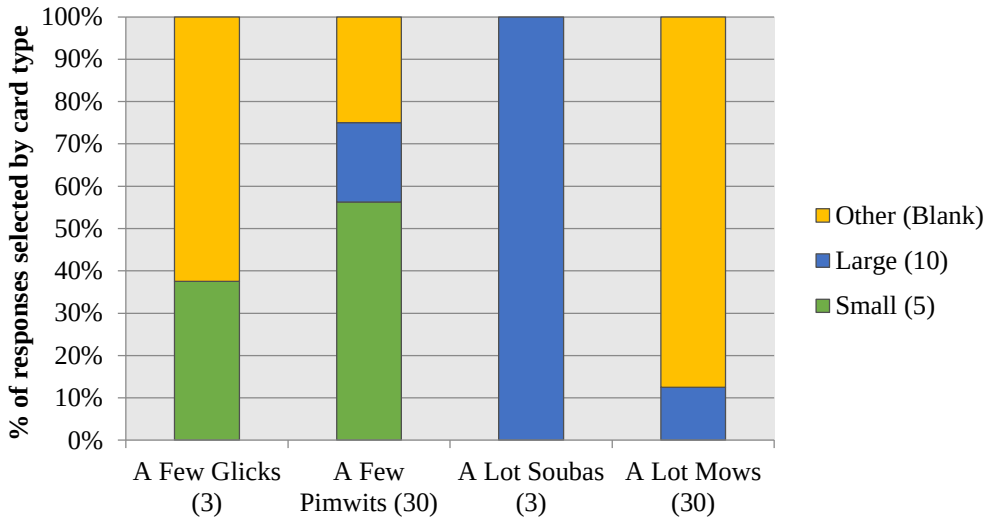
First we will discuss the overall pattern of results of the two Quantity Conditions combined, and then will discuss findings from each condition. The results from both Quantity Conditions can been seen in Figure 2.

Figure 2. *Percent card selection responses given for each trial in the (a) 5-40 Condition, and (b) 3-30 Condition. The numbers on the x-axis reflect the quantities presented in the distribution trials.*

(a)



(b)



A Few Trials

In the "a few" trials the two hypotheses make opposite predictions about what should happen. In the 5-40 Condition the Fixed Quantities Hypothesis predicts that participants should reject the small card (10 items) in all trials, regardless of context, and should instead pick the blank card. This prediction is based on the assumption that "a few" approximately picks out quantities between 3-5, thus, the smallest answer card in this Condition, which displays 10 items, shows too many items to be the correct quantity for "a few" and should be rejected. Additionally it predicts that if people pick the blank card consistently, that they should also consistently report the quantity on the blank card to be within the range of "a few". For the 3-30 Condition, it predicts that participants should pick the small card (5 items) regardless of context as the quantity falls within the range of "a few". Instead, the Relative Quantities Hypothesis predicts that participants will be more willing to accept the small card (10 items or 5 items) in the larger average trials (distributional averages of 40 or 30), because relative to the distributional information provided to them, the

quantities visible on the small cards are comparatively small quantities that could be plausible quantities for "a few". Additionally, it predicts that participants will reject the small card in both of the smaller average (distributional averages of 5 or 3) trials, as the quantities presented on the small cards are in fact greater than the average. Essentially it makes the prediction that there should be an effect of context.

In the "a few" trials we found that more participants rejected the available quantities and selected the blank card in the smaller quantity trials (when the average was 3 or 5), than in the larger quantity trials (when the average was 30 or 40). The difference between responses on the two sets of trials was significant ($p < .001$; FET). Specifically, we find that in the smaller quantity trials participants selected the blank card 78.13% of the time, versus 25% of the time in the larger quantity trials. In the larger quantity trials we find that participants accept the small quantity card 65.63% of the time, versus 21.88% of the time in the smaller quantity trials.

When looking at each condition separately we found that the effect was stronger in the 5-40 Condition ($p < .001$; FET) than in the 3-30 Condition ($p = .05$; FET). Despite this difference, we still overall find an effect of context. These findings provide evidence that even in with minimal knowledge of a novel context the effect of context is still present when people are interpreting the meaning of the term "a few". Given that we find this effect of context in novel situations where participants are given the option to select the blank card and choose a quantity within the fixed range of a few, and yet do not in the large quantity trials provides strong evidence for the Relative Quantities Hypothesis.

A Lot Trials

In the "a lot" trials, the two hypotheses again make opposite predictions. The Fixed Quantities Hypothesis predicts that participants should consistently pick the same quantity for "a

lot" across all trials in both conditions. In opposition, the Relative Quantities Hypothesis predicts that participants will be more willing to accept the large card (10 items or 20 items) in the smaller average trials (3 or 5), and reject the larger card in the larger average (30 or 40) trials. Essentially it makes the prediction that there should be an effect of context.

Similar to the findings of the "a few" trials we found an effect of context in the "a lot" trials. Mainly, we found that participants were more willing to select the blank card on the larger quantity trials than on the smaller quantity trials ($p < .001$; FET). We found that participants selected the blank card 71.88% of the time in the large quantity trials, versus only 3% of the time in the smaller quantity trials. Instead, in the small quantity trials participants were more likely to select the larger quantity answer card (87.5%, versus 28. 13% in the larger quantity trials). When looking at each condition separately we found that the effect was stronger in the 3-30 Condition ($p < .0001$; FET) than in the 5-40 Condition ($p < .001$; FET). Given that these results reflect the fact that the contextual information influences the quantities people are willing to accept for "a lot", these findings could be said to also support the Relative Definition Hypothesis. However, it could be noted that it is actually much harder to make an assumption about what the fixed range for "a lot" might be. While, it may be difficult to ascertain a prototypical lower bound for "a lot", the results indicate that 100% of participants in the 3-30 Condition were willing to accept 10 objects as being "a lot" in the small quantity trials, and 18.75% were willing to accept 10 objects as being "a lot" in the small quantity trials in the 5-40 Condition, yet many of the same participants reject 10 as being "a lot" in the large quantity trials. If the Fixed Quantities Hypothesis is correct, then it is difficult to explain this particular pattern of results, instead it would have predicted that the minimum quantity for "a lot" remain consistent across contexts.

The findings from this study provide some evidence that even with minimal contextual information adults use a strategy that appeals to contextual information when determining the meaning of terms such as "a few" and a "a lot" supporting the Relative Quantities Hypothesis. One potential limitation of this study is perhaps the fact that the contextual information was uniform. Given that all of the information was identical, it is possible that participants rationalized that it was impossible or highly unlikely that anyone owned quantities of the novel objects within the rages of "a few" and "a lot" in the trials where they predominantly chose the blank card. While we believe that the findings from Experiment 1 provides evidence that this conclusion is unlikely, further studies are planned that use varied distributions that include in-range quantities, while still having larger or smaller overall averages, which could eliminate this confound.

Additionally it would be interesting to run this study with two possible methodological changes. In one version of this task, it might be interesting to run trials looking at the terms "a couple" and "dozens" as controls. We would thus expect no effect of context for these terms, and for example, in the case of "a couple" we would predict that when 2 is not an answer option all participants should select the blank card as the correct answer. A second potential methodological change would be to split participants into two groups, one that receives distributional information, and one that does not, and instead of presenting them with answer cards, simply ask participants to estimate how many of the novel items the test subject has (e.g., "Mike has *a few* glicks. How many glicks do you think he has?"). We might expect that the estimated quantities remain fairly constant for each quantity term in the no distributional information group across contexts, and to vary by context for "a few" and "a lot" in the distributional information group.

General Discussion

Some quantity words, such as numerals pick out specific quantities in the world, e.g. the word “two” in the sentence “two birds flew over the fence” picks out exactly two individual birds, other words such as “a few” and “a lot” are much less restricted in the quantities they pick out. In a pilot study we determined that not only are the terms “a few” and “a lot” less restricted in the quantities that they pick out than “two”, but it would appear that contextual knowledge plays a role in the quantities picked out by each term. Specifically, if the term “a few” was used in a sentence to refer to some quantity of things that people typically associate with large quantities (such as the number of friends people typically have on Facebook) we saw that people estimated that the term reflected larger quantities than when used in a sentence to refer to some quantity of things people typically associate with smaller quantities (such as the number of houses owned by an individual). The experiments in this paper attempted to address how this could be possible, and what tools people might be accessing when they attempt to interpret these kinds of terms.

We suggested that there were two potential hypotheses that could explain the role of context in interpreting these terms. The first hypothesis, the Fixed Quantities Hypothesis, suggested that the role of context was not inherently part of the meaning, or semantics, of the terms themselves, but instead was made out of social considerations of the listener. This hypothesis presupposes that these terms have exact meanings, and pick out specific quantity ranges in the world (akin to how numerals pick out specific quantities). It is also crutched on the belief that people are being generous in their interpretation of the terms, meaning that while they might think that a term is being incorrectly used, they can be flexible in their interpretation and make an assumption about what the speaker likely meant.

The second hypothesis, the Relative Quantities Hypothesis, in contrast, assumes that the role of context is built into the very meaning of the terms. This hypothesis is grounded in theories on the semantics of relative gradable adjectives, e.g., tall, short, where the meaning of the term is dependent on the noun. For example, what counts as "tall" for a tree is very different than what height qualifies as "tall" for a cup. Basically, relative gradable terms require accessing the relevant referent set to get some information about how to interpret the term, and thus referencing contextual information is built into the very definition of the terms.

These two hypotheses make very different predictions for each of the three experiments we conducted in this paper. Each experiment attempted to tease apart these two hypotheses and determine the exact role of context in the interpretation of these terms. The Relative Quantities Hypothesis predicted that the reason we find an effect of context in the usage of the quantity terms "a few" and "a lot" is simply because the ranges that they quantify is dependent on the context they are quantifying over. Alternatively, the Fixed Quantities Hypothesis suggested that the reason why we found variability in the quantities given for "a few" and "a lot" in our pilot study was due to participants finding their uses to be infelicitous or improbably for certain contexts, and therefore were being generous in their interpretations of the terms. Experiment 1 investigated this hypothesis by asking participants to rate whether or not they thought quantities within the range of "a few" were im/probable for a variety of contexts. The findings suggested that mostly participants did not find small quantities implausible, even for the larger average contexts. This evidence goes against the assumptions made by the Fixed Quantities Hypothesis.

In Experiment 2, we further tested this hypothesis by attempting to reduce the pragmatic pressure to accept uses of the terms that might be infelicitous. The task itself requested that participants to rate how acceptable they felt the use of each term was in different contexts, and to

describe different quantities. Given that participants were provided with the opportunity to say that the use of a term was incorrect the Fixed Quantities Hypothesis predicted that the results for "a few" and "a lot" should not have varied based on context, and instead should have been more like the results for "dozens" or "a couple". Instead we still found the effect of context. Finally in Experiment 3, we specifically tested out the Relative Quantities Hypothesis by giving participants limited exposure to contextual information for novel stimuli, and yet still found an effect of context in the interpretation of the terms "a few" and "a lot".

The evidence provided by Experiments 1 through 3, give strong evidence for the Relative Quantities Hypothesis, indicating that there are a class of quantifiers whose meanings vary, and are context-dependent. Additionally, evidence from Experiment 2 points to the possibility that there are a group of quantifiers that do have fixed definitions. What we found in Experiment 2, was evidence that there are at least two quantifiers whose meanings do not really vary, and instead appear to be fixed. These findings leave open a number of interesting questions to explore. Future studies should aim to explore the exact semantic nature of these terms, and draw parallels between their semantic structure and the semantic structure of relative and fixed gradable adjectives.

Future experimental work should address what specific contextual evidence is used to influence the meaning of quantifiers. During the course of this paper we found some evidence that world knowledge of average information could be related to how people make estimations about the meaning of "a few", but the findings of this paper cannot say with certainty that average information is all of the contextual information being used to interpret these relative definition quantity terms. Additionally, how exactly average information might be influencing the meanings of these terms is still an open question. One potential answer comes from research that has shown that speaker-specific variability can play a role in how people resolve the meanings of quantity

terms (Yildirim, Degen, Jaeger, & Tanenhaus, 2013). These findings suggest that people are making these adaptations by making judgements using statistical information about the speaker's usage of a term. In the case of this particular study, participants modified their reference set for the meaning of terms like "some" and "many" to match the information made available to them. Thus it seems as if people can rapidly make judgements about the meanings of quantity terms by accessing distributional usage information. Thus, perhaps people might be using similar statistical computations to make judgements about relative quantity terms. If it is the case that they are like relative gradable adjectives, then there is some evidence that supports this hypothesis (see: Schmidt, Goodman, Barner, & Tanenbaum, 2009).

Future research is planned to examine this potential role of distributional statistical information on the meanings of these relative quantity terms. We plan to conduct a study similar to Experiment 3 using novel stimuli, and introducing participants to a more diverse sample of information (e.g., wider distributions), and to observe which factors predict the shifts in each terms' meaning. Additionally, we would like to extend this research to look at how young children learn to incorporate contextual information, and to draw parallels or find differences in the ways children learn to accommodate for reference set information in interpreting relative quantity terms, versus relative gradable adjectives.

References

Brown, P., & Levinson, S. C. (1987). *Politeness: Some Universals in Language Usage.* Cambridge: Cambridge University Press.

Deacon, T. W. (1998). *The Symbolic Species: The Co-evolution of Language and the Brain.* New York, NY: W. W. Norton & Company.

Kennedy, C. (2007). Vagueness and Grammar: The Semantics of Relative and Absolute Gradable Adjectives. *Linguistics and Philosophy, 30*, 1-45.

Lasersohn, P. (1999). Pragmatc Halos. *Language, 75*, 522-551.

Leech, G. N. (1983). *Principles of Pragmatics.* New York: Longman Group Limited .

Pinker, S. (1994). *The Language Instinct.* New York, NY: William Morrow and Company.

Schmidt, L., Goodman, N. D., Barner, D., & Tenenbaum, J. B. (2009). How tall is tall? Compositionality, statistics, and gradable adjectives. In N. A. Taatgen, & H. van Rijn (Ed.), *Proceedings of the 31th Annual Conference of the Cognitive Science Society* (pp. 3151-3156). Austin, TX: Cognitive Science Society.

Sperber, D., & Wilson, D. (1986). Loose Talk. *Proceedings of the Aristotelian Society, 86*, 153-171.

Syrett, K., Kennedy, C., & Lidz, J. (2010). Meaning and Context in Children's Understanding. *Journal of Semantics, 27*, 1-35.

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *43*(3), 631-650.

Yildirim, I., Degen, J., Jaeger, T. F., & Tanenhaus, M. K. (2013). Linguistic variability and adaptation in quantifier meanings. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Ed.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 3835-3840). Austin, TX: Cognitive Science Society.