Why is A Few Sometimes A Lot?

Amanda Pogue (amanda.pogue@uwaterloo.ca)
University of Waterloo, Department of Psychology,
200 University Avenue West, Waterloo, ON N2L 3G1 Canada

Adel Jalabi (ajalabi@uwaterloo.ca)
University of Waterloo, Department of Psychology,
200 University Avenue West, Waterloo, ON N2L 3G1 Canada

Mathieu Le Corre (mlecorre@uwaterloo.ca)
University of Waterloo, Department of Psychology,
200 University Avenue West, Waterloo, ON N2L 3G1 Canada

Abstract

It is not surprising to find that the quantity picked out by terms like *a few* and *a lot* is context dependent. We can easily accept that *a few* books might be 10 books, yet *a lot* of smartphones might only be 4 smartphones. The current paper posits that there are two hypotheses that can explain can explain this context dependency: the Definite Number Hypothesis (DNH), and the Gradable Quantifier Hypothesis (GQH). The DNH suggests that the term *a few* corresponds to a definite range of values, and may pick out a larger quantity only if the range seems implausible for the given context. The GQH suggests that context-dependency is actually built into the meaning of *a few*. Experiment 1 supports the intuition that there is variability in the quantity that *a few* picks out based on context. The findings of Experiments 2 and 3 support the Gradable Quantifier Hypothesis.

Keywords: quantity-judgment; quantifiers; semantics; psycholinguistics; context

Introduction

When we consider terms such as *a few* and *a lot*, we intuitively know that relative to each other *a few* will always be smaller than *a lot*; however, what does this tell us about how we should evaluate the quantity that they represent across different sets in the world? One potential strategy is to assume that *a few* is always a small quantity and that *a lot* is always a large quantity. However, this strategy may lead us astray. For example, in Experiment 1 we find that 19 English-speaking adults suggest that *a few* friends on Facebook® is about 76 friends, whereas *a lot* of houses owned by a single individual is approximately 4.3 houses. Given this finding, it seems highly unlikely that the aforementioned strategy is the best for describing how we evaluate these terms in the wild. How then might we be dealing with these terms?

There is no doubt that we must be using some sort of contextual information to guide our interpretations. But how do we do this? We constructed and consider two hypotheses. On the Definite Number Hypothesis, the meaning of *a few* corresponds to a definite range of values – namely, small subitizable quantities (2 to 4 or 5). It was predicted that this range might be determined by comparison to other similar quantity terms using pragmatics (see: Grice, 1989; Barner, Brooks, & Bale, 2011). Given that adults could easily select other terms such as *a couple* or *a pair* to represent 2 items, and terms such as *several* or *a handful* to represent slightly larger quantities, it was predicted that people could interpret the term *a few* with relation to these quantity terms, and therefore give *a few* a definite range of 3 to approximately 5. However, when speakers use *a few Xs* in contexts where 3 to 5 Xs is an extremely atypical quantity of Xs (say, Facebook friends), hearers interpret what the speakers have said by making a post-hoc adjustment from the definite meaning of *a few* (i.e., 3 to 5) to the smallest plausible quantity, say about 70 for Facebook friends. On the Gradable Quantifier Hypothesis, the meaning of *a few* is inherently context-sensitive, much like the meaning of gradable adjectives like *small*, or *tall*. In other words, the range of quantities corresponding to *a few* cannot be fully specified in the abstract; it can only be specified given some particular context. For example, there is no sense in asking what degree of height corresponds to *small* in general. Rather, one must know what type of individual is being measured, and what is the typical height for that type of individual (see: Kennedy, 2007; Syrett, Kennedy, & Lidz, 2010). The same may be true for the range of quantities that correspond to *a few* (i.e., one may have to know what type of individual is being quantified, and what is the typical quantity in which such individuals are found).

Both hypotheses are consistent with the fact that the numerical value associated with *a few* varies with context. Therefore, that fact alone cannot adjudicate between them. However, they attribute this variability to different factors. According to the Definite Number Hypothesis, the variability of *a few* is a function of whether the quantities in the range of its meaning (i.e., 3 to 5) are plausible; if they are not, then one adjusts to the smallest plausible value for the context at hand. In contrast, according to the Gradable Quantifier Hypothesis, the likelihood of finding 3 to 5 individuals of a given type matters not at all. Rather, all that matters is the typical quantity in which these individuals cluster. In the current paper, we test these hypotheses in two ways. First, in Experiment 1 we confirm the intuition that there is variability in the quantities associated with *a few*. In Experiment 2, we ask whether one observes variability in the quantities associated with *a few*, even in contexts where

3 to 5 are highly plausible quantities. Finally, in Experiment 3, we ask whether speakers associate *a few* with various quantities when they have information about typical quantities, but no information concerning whether 3 to 5 are plausible quantities. Whereas the Definite Number Hypothesis predicts that there should be no variability in the quantities associated with *a few* in Experiments 2 and 3, the Gradable Quantifier Hypothesis predicts the opposite pattern of results.

Experiment 1

In Experiment 1 we aimed to look at what quantities English-speaking adults attribute to the terms *a few* and *a lot* with regards to common every day scenarios. Participants were asked to estimate values of *a few* and *a lot* when they were used in context of various common real world situations. It was expected that participants' responses would be affected by context (the estimated average for each situation), such that the overall averages given for both *a few* and *a lot* would vary for each scenario based on the average.

Methods

Participants Nineteen fluent English-speaking adults (M = 20.7 years old, 15 females) were recruited from the University of Waterloo. Participants received partial course credit for participating in the study.

Materials and Procedure Participants were asked to fill out a questionnaire where they were required to estimate the values associated with a series of one-sentence scenarios. Twenty different scenarios were constructed. Fourteen of the scenarios were about types of things individuals might own / have (dogs, cats, cars, computers, bicycles, books, children, guitars, smartphones, houses, DVDs, shirts, pennies, and TVs), and the remaining 6 items were about things and events in the world (trees in a park, apples picked in an outing, goals scored in a season, emails received over a week, friends on Facebook, and photos tagged online). Half of the items were expected *a priori* to have very small averages (Small Quantity Items), and the other half were expected to have relatively higher reported averages (Large Quantity Items). Each of the scenarios were described by one sentence containing the target term *a few* (1), or *a lot* (2), and were paired with a question asking the participant to evaluate how many items were described in each scenario. Additionally, the questionnaire included items asking participants to evaluate what the average might be for each scenario (3).

(1) Dan has a few friends on Facebook. How many friends do you think he has on Facebook?
(2) Dan has a lot of friends on Facebook. How many friends do you think he has on Facebook?
(3) What is the average number of friends that people might have on Facebook?

The questionnaire contained all 60 test items. Each of the terms (a few, a lot, and average) were presented in separate blocks. The order of the scenarios was counterbalanced across blocks, and the order of the blocks was counterbalanced across participants. Each block appeared on a separate page.

Results

Participants were elicited to pick a number to answer each of the test items. Since they were not given specific numbers to choose from their responses were fairly variable. As a result, before conducting any analyses we ran a recursive outlier detector on the responses made by each participant per test item, 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 responses. We then conducted a 2x2 repeated measures ANOVA with Quantifier (a few, a lot, and average) and Test Item as the repeated measures, and Subject Gender and Order as the between subjects measures. The results indicated an effect of Test Item ($F(19,19) = 3.055$, $p < .001$), and an interaction of Test Item and Quantifier ($F(38,38) = 2.655$, $p < .03$). Indicating that the results differ based on the individual test items, and that the quantifier used plays a role in the results for each test item. There was no effect of Quantifier, Subject Gender, or Order, nor any other significant interactions ($ps > .1$).

We conducted several planned comparisons to investigate the interaction. As predicted there was a significant difference between the Small Quantity Items (M = 1.48), and the Large Quantity Items (M = 73.57) when the participants were asked to estimate the means for the items ($t(18) = 9.89$, $p < .001$). Similarly, we found a difference between the Small Quantity Items (M = 2.26), and the Large Quantity Items (M = 20.63) when the participants were asked to estimate what was meant by *a few* ($t(18) = 6.4$, $p < .001$), and a difference between the Small Quantity Items (M = 4.96), and the Large Quantity Items (M = 264.34) when the participants were asked to estimate what was meant by *a lot* ($t(18) = 6.69$, $p < .001$).

Experiment 2

Both the Definite Number Hypothesis and the Gradable Quantifier Hypothesis can explain the variability found in Experiment 1. In Experiment 2 we ask if we can find contexts where having 3 to 5 items is highly plausible, but where people interpret *a few* as picking out quantities larger than 5.

Experiment 2 provided participants with contexts where the use of *a few* could plausibly be used felicitously to describe 3 to 5 items, or could also be used to describe a potentially larger quantity based on the participants' world knowledge. Participants were also asked to give a plausibility rating for scenarios such as in (4) which incorporated the numbers 3 to 5 rather than using the terms *a few* or *a lot.* The Definite Number Hypothesis predicts a very low plausibility rating for all contexts where *a few*

picks out a quantity greater than 5. Alternatively, the Gradable Number Hypothesis is more agnostic about the role of plausibility, but might predict that the all of the contexts where *a few* picks out quantities larger than 5 should be considered plausible.

(4) Martha has four friends on Facebook.

Methods

Participants Forty-four fluent English-speaking adults (M = 20.4 years old, 25 females) were recruited from the University of Waterloo. Participants received partial course credit for participating in the study.

Materials and Procedure Participants were asked to fill out a questionnaire where they were required to estimate the values associated with a series of one-sentence scenarios (Estimation Survey). Seven new scenarios were constructed. These scenarios included: silver cars in a parking lot, trips 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. These items were chosen because it was predicted by the authors that they were contexts where *a few* could pick out a quantity larger than 5, but that they could plausibly be used in scenarios with quantities between 3 to 5. Unlike in Experiment 1, each of the scenarios was used in just a single trial that used the target term *a few*. The order of the scenarios was counterbalanced across participants.

Each participant completed a second survey following the Estimation Survey. The second survey (Plausibility Survey) included the 7 new scenarios introduced in Experiment 2, and 9 of the items from Experiment 1 (specifically items where *a few* picked out larger quantities). Each context was turned into a one sentence scenario which incorporated a number between 3 to 5 (e.g., "There are three silver cars parked in the UW parking lot today."). Participants were asked to rate the plausibility for each of these sentences on a scale of 1 to 5, where 1 is highly implausible, and 5 is very plausible. The Plausibility Survey was always completed after the Estimation Survey, and was always presented on a separate page. The order of the scenarios was randomized between old and new items, and was counterbalanced across participants.

Results

Table 1 reports the average plausibility ratings for all the test items rated in Experiment 2, and the average quantity associated with *a few* for each of these items.[1] 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). Of the items analyzed, the number of friends on Facebook was the only one where subjects associated a large quantity with *a few*, but where quantities between 3 and 5 were judged as having low plausibility. This provides evidence that *a few* is genuinely gradable.

Table 1: Mean responses given for "a few X" in Experiment 1 (noted by*) and Experiment 2, and the mean plausibility rating for each corresponding trial in Experiment 2

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	9.85
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*

Experiment 3

Experiments 1 and 2 provide evidence that adults use context to adjust their interpretations of *a few* and *a lot*. Experiment 2 further showed that it is unlikely that this effect is as a result of the participants interpreting the use of the term *a few* as being infelicitous.

Experiment 3 provides further evidence that *a few* must be gradable, as predicted by the Gradable Quantifier Hypothesis. In Experiment 3 participants are told about novel objects that people from Southern Mexico like to collect. It was assumed that the participants had little to no experience with people from Southern Mexico or the things that they like to collect, thus, the participants could not use any of their prior knowledge to make inferences. Instead, we provided the participants with information about the average number of each of the objects owned by individuals, and then asked them to estimate how many items are picked out by *a few* or *a lot* for each of the novel items. Since the participants have no prior experience with the objects, and since we did not provide them with any information about the plausibility of individuals only owning 3 to 5 of the items, participants must rely on the information provided to them.

[1] Note: Similar to Experiment 1 participants were asked to make personal judgments and the responses varied greatly. A test for outliers (Van Selst & Jolicoeur, 1994) suggested that 6.82% of responses be removed due to being significant outliers. The data reported in Table 1 excludes the outliers in Experiment 1 and Experiment 2.

The Definite Number Hypothesis predicts that in Experiment 3, all participants would say that *a few* is 3 to 5 items regardless of the provided average information. Conversely, the Gradable Quantifier Hypothesis predicts that the number that *a few* picks out in Experiment 3 should vary depending on the context.

Methods

Participants Sixteen fluent English-speaking adults (M = 21.6 years old, 10 females) were recruited from the University of Waterloo. Participants received partial course credit for participating in the study.

Materials and Procedure Participants were seated at a computer table beside an experimenter. The study was composed of three parts: two training tasks, and the main task. First, the participants were introduced to the format of the task. They were instructed that they would be told some short stories. Accompanying these stories would be pictures displayed on the computer screen. They were instructed that when the stories were completed that they would be asked questions about the stories. In order to respond to these questions the participants were asked to make a selection between three cards visible on the screen. Participants would indicate the answer they thought was correct. Additionally, they were told that only two of the cards would have visible answers, and if they did not think that the correct answer was on the visible cards that they should choose the blank card.

In the first training task participants were given three trials where they were asked to locate specific items (e.g., toy car, teddy bear, 2 monkeys) using the three card response system, and received feedback on their choices.

In the second training task participants were introduced to a picture of a novel object, and were told that it was a kind of object that people in Southern Mexico like to collect. They were then shown a slide that looked like the top left box in Figure 1, and told (5). For the second training task the number of items shown on the screen was either 2 or 3. They heard three stories with three different individuals, each with the same number of items. Participants were then asked to guess how many of the novel objects a fourth individual might own, and were given the choice between the same number of items from the previous slides (2 or 3), and another amount (1 or 5). The purpose of this training trial was to determine if the participants could abstract average information from three short stories. There were two trials in the second training task. No feedback was given on these trials.

(5) This is Sarah, and she has this many Xs.

The test trials were identical to the second training task, except the design of the question response slide. On each trial participants were told stories about a novel object that people in Southern Mexico like to collect. They were told about three individuals, and were shown how many objects each of those individuals owned. Each participant heard 4 stories, of which half of the stories were about characters who owned an average of 5 of the novel stimulus item (Small Quantity Context), and the other half were about characters that owned an average 40 of the novel stimulus item (Large Quantity Context). The novel stimuli items were always presented inside of a box shaped like a rotated card. The size of the box, and the size of the items remained constant throughout the story slides, and in the question response slide. It was assumed that visual cues such as the density of the items should provide evidence to indicate a difference between the quantities presented in the story slides, and those on the question response slides.

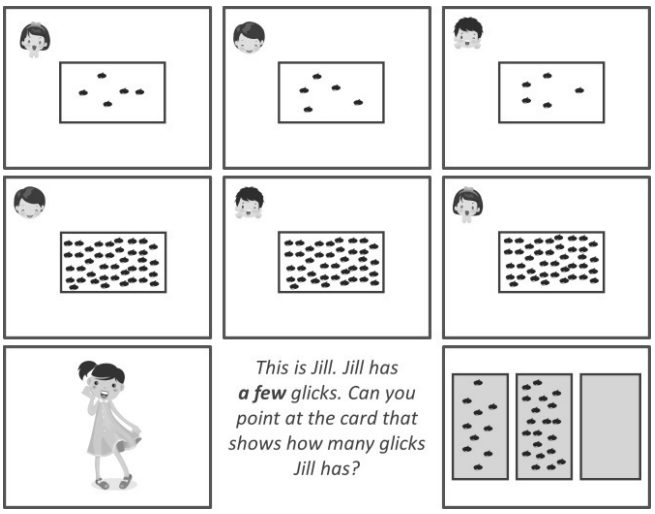

Figure 1: Example of Experiment 3 design. Row 1: an example of the slides that correspond with each of the three stories in the Small Quantity Context. Row 2: example of the slides that correspond with each of the stories in the Large Quantity Context. Row 3: the 4th character, the test question, and question response slide.

After hearing about how many items three characters had, participants were introduced to a fourth character and were told that the character "has <*a few* / *a lot*> of Xs." After hearing this statement, participants were shown the question response slide with the three cards, and were asked to estimate how many of the novel objects they thought that the fourth character had. Participants had the option to pick between three cards: 10, 20, and blank, for each of the stories. If the participant chose the blank card they were asked to estimate how many items should be on the card. Each term was used once per context, and the order of the terms and contexts were counterbalanced across participants.

Results

Figure 2 reports the percentage of response by card choice for each of the conditions in the task. As predicted by the Definite Number Hypothesis 93.75% of the participants chose the blank card in the Small Quantity Condition in the

a few trial, and reported that the quantity picked out for *a few* in this trial should be ~3.9. However, contrary to the Definite Number Hypothesis only 25% of the participants picked the blank card in the Large Quantity Condition, reporting that the quantity picked out for *a few* in this trial should be ~3.3. The remaining 6.25% and 75% respectively chose the small quantity card indicating that the quantity picked out for *a few* for these trials should be 10. Participants were significantly more likely to pick the blank card in the Small Quantity Condition for *a few* than in the Large Quantity Condition ($\chi^2(1, N = 16) = 9.091$, $p < .01$). Additionally, participants were significantly more likely to pick the blank card in the Large Quantity Condition for *a lot* than in the Small Quantity Condition ($\chi^2(1, N = 16) = 4.900$, $p < .03$).

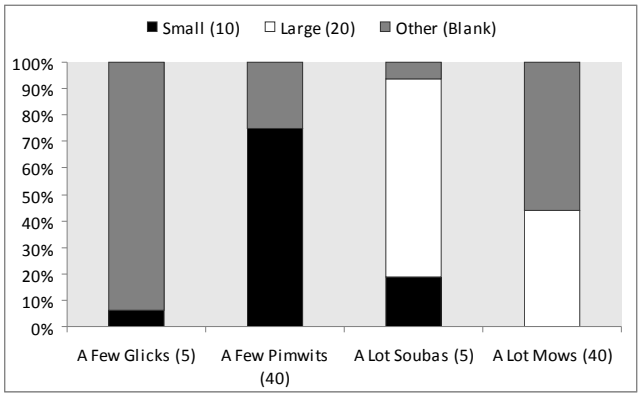

Figure 2: Percent per response type for each trial in Experiment 3.

General Discussion

The results of the experiments presented in this paper provide direct evidence that there is variability in the quantities picked out by the term *a few*, and that this variability can be explained by context. This finding is highly unsurprising – no one would argue that context plays a role in the interpretation of all quantifiers. However, we set out to investigate more than just the effect that context has on the quantities picked out by *a few*, and endeavored to figure out how context is having this effect. We speculated that there are two hypotheses that could explain the role that context is playing for *a few*: 1) the Definite Number Hypothesis, and 2) the Gradable Quantifier Hypothesis.

According to the Definite Number Hypothesis, the variability of *a few* is a dependent on whether the quantities in the range of its meaning (i.e., 3 to 5) are plausible. If the quantities are not within the range of its meaning (like in the case of Facebook friends), then they must be adjusted to the smallest plausible value for the context at hand. Alternatively, according to the Gradable Quantifier Hypothesis, the likelihood of finding 3 to 5 individuals of a given type is not what matters, but rather the typical quantity in which these individuals cluster. These two hypotheses made separate predictions regarding the outcomes of Experiments 2 and 3 in this paper.

In Experiment 2, the Definite Number Hypothesis predicted that participants would rate all of the contexts that for which quantities plausibly fall within its range as being highly plausible, and those for which quantities do not plausibly fall within its range as being implausible. The Gradable Quantifier Hypothesis makes no such prediction, and is not constrained in its meaning. Thus, it should have predicted that the plausibility ratings given by participants should not depend on the plausibility of the quantities falling within a specific range. The findings of Experiment 2 go against the predictions of Definite Number Hypothesis, as there was evidence for several items that do not fall within the range of *a few*'s definite meaning, yet were rated as being plausible scenarios. Though the prediction made by the Definite Number Hypothesis did hold true for the context involving friends on Facebook, but only for that one context. The results of Experiment 2 consequently seem to imply that people do not expect *a few* to have a definite range that requires post-hoc adjustments for larger contexts.

In order to further test these hypotheses, in Experiment 3 the only context information made available to participants was average information. Participants could not rely on their prior knowledge with the novel items to determine whether the definite range of quantities picked out by *a few* in the Definite Number Hypothesis is plausible for the novel contexts. Thus, the Definite Number Hypothesis predicted that participants should indicate that the quantity picked out by *a few* in all contexts should fall within the range of meaning (i.e., 3 to 5). On the other hand, the Gradable Quantifier Hypothesis predicted that *a few* would pick out a quantity relative to the typical quantity (i.e., average) for each context. Consequently, it would predict that the responses for the Small Quantity Condition would be different from the responses for the Large Quantity Condition. The results of Experiment 3 support the hypothesis predicted by the Gradable Quantifier Hypothesis, in that there was a significant difference in the responses given for the different context scenarios. Conversely the Definite Number Hypothesis only correctly predicted the responses for the Small Quantity Condition.

One potential limitation of Experiment 3 is that it is possible that while participants were not explicitly supplied with information that confirmed or denied the probability of 3 to 5 items being plausible quantities in the novel contexts, the participants may have inferred that the 3 to 5 would be implausible in the Large Context Condition. While this may be the case, when the results of Experiment 3 are considered with respect to the results of Experiment 2, which suggested that that we still see variability in the quantities picked out by *a few* despite 3 to 5 being plausible quantities for the given contexts, the Gradable Quantifier Hypothesis still provides a better explanation for the data.

The future direction of this research aims to investigate several questions left open by the findings in this paper. First, if the Gradable Quantifier Hypothesis is true, how

does context tell us what quantity is picked out by *a few*? The answer to this question can perhaps be found by determining the semantic form of *a few* and other gradable quantifiers with respect to the form of gradable adjectives. Secondly, it would be interesting to investigate whether all quantifiers are gradable like *a few,* or if it is the case that only some quantifiers are gradable in this way; it would be interesting to study what makes a specific quantifier gradable. Finally, the structure of gradable quantifiers seems like it would be very difficult to acquire. Given that both *a few* and *a lot* can be used in variable contexts to mean either a small quantity, or a relatively large quantity, how do children learn that what is important about gradable quantifiers is the context in which they are used, and not a definite number range?

The results of the present study support the Gradable Quantifier Hypothesis, suggesting that context-dependency is built into the meaning of terms such as *a few*, and rejects the Definite Number Hypothesis. This finding opens up a host of interesting questions about the gradability of other terms, and the acquisition of gradable quantifiers. Current research in the authors' lab aims to answer these questions with future and ongoing studies.

Acknowledgments

Thanks to the members of the Developmental Science Lab and the members of the Developmental division in the Department of Psychology at the University of Waterloo for the useful feedback. A.P. would also like to thank Christie Haskell, Jonathan Fugelsang, and Hugh Rabagliati for the statistics advice. This research was supported by a grant to M.L. from the Natural Sciences and Engineering Research Council of Canada.

References

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition, 188,* 87-96.

Grice, P. (1989). *Studies in the way of words.* Cambridge, MA: Harvard University Press.

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

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

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size and skew on outlier elimination. *Quarterly Journal of Experimental Psychology, 47A,* 631--650.