# *Surprisingly*: Marker of Surprise Readings or Intensifier?

**Anthea Schöller and Michael Franke**
(anthea.schoeller@uni-tuebingen.de, mchfranke@gmail.com)
Department of Linguistics, Wilhelmstraße 19
72074 Tübingen, Germany

## Abstract

We investigate the influence of the adverb *surprisingly* on the meaning of the quantity words *few* and *many*, which themselves have been associated with a reading expressing surprise. To learn about the meaning contribution of "surprise", we compare *surprisingly* with the intensifier *incredibly* and a *compared to* phrase explicitly marking surprise. Based on an empirical measure of subjects' expectations about everyday events, a Bayesian model uses data from a sentence judgment task to infer likely levels of surprise associated with the different constructions of interest.

**Keywords:** intensifier, surprise, computational modeling, *few*, *many*, *surprisingly*

## Introduction

A long tradition in psychology has acknowledged the role of prior expectations in the use of vague and context-dependent expressions like *tall, heavy, few* and *many* (e.g. Clark, 1991; Sanford, Moxey, & Paterson, 1994). Fernando and Kamp (1996) spell out a semantic theory which makes the truth conditions of *few* and *many* dependent on prior expectations. So-called "cardinal surprise readings" convey that a cardinality is smaller or greater than what is expected for the situation:

(1) For a man from the US, Chris saw few/many movies last year. ⇝ Chris saw less/more movies than expected for a US male.

Such a surprise-based account raises interesting questions. First, how can expected cardinalities be distinguished from surprising ones? Fernando and Kamp (1996) stipulate that the lexical meanings of *few* and *many* comprise contextually-stable thresholds $\theta_{\text{few}}$ and $\theta_{\text{many}}$ which operate on a contextually-variable representation of *a priori* expectations. Second, if sentences with *few* and *many* express that a cardinality is surprising anyway, are they different from sentences in which the surprise element is overtly marked? The surprise reading can be made salient by a *compared to* phrase (2) or by modifying *few* and *many* with an adverb like *surprisingly* (3).

(2) Compared to what you would expect for a man from the US, Chris saw few / many movies last year.

(3) For a man from the US, Chris saw surprisingly few / many movies last year.

If surprise were the single factor which determines truth conditions of the cardinal surprise reading, we should not find a meaning difference between (1) and the overtly marked surprise reading in (2) and (3). Alternatively, it could be hypothesized that *surprisingly* in (3) acts not as a marker of surprise but as an intensifier, yielding a higher $\theta_{\text{surpr. many}}$ than $\theta_{\text{many}}$ and a lower $\theta_{\text{surpr. few}}$ than $\theta_{\text{few}}$. The pragmatic theory of intensifiers by Bennett and Goodman (2015) would predict that *surprisingly* has very similar effects to *incredibly* (see below).

We set out to experimentally test the influence of the modifiers *surprisingly, incredibly* and *compared to* on the threshold values predicted for Fernando and Kamp (1996)'s surprise readings of *few* and *many*. We employ linear mixed effects regression to compare judgment data and a computational model to infer said thresholds from our data.

## A Surprise-based Semantics for *few* and *many*

Partee (1989) characterized cardinal *many* as describing cardinalities which are *at least* $x_{\text{min}}$, where $x_{\text{min}}$ is a large number, and *few* as describing cardinalities which are *at most* $x_{\text{max}}$, where $x_{\text{max}}$ is a small number, see (4).

(4) Cardinal reading of "Few/Many As are B"
*Few*: $|A \cap B| \leq x_{\text{max}}$ *Many*: $|A \cap B| \geq x_{\text{min}}$

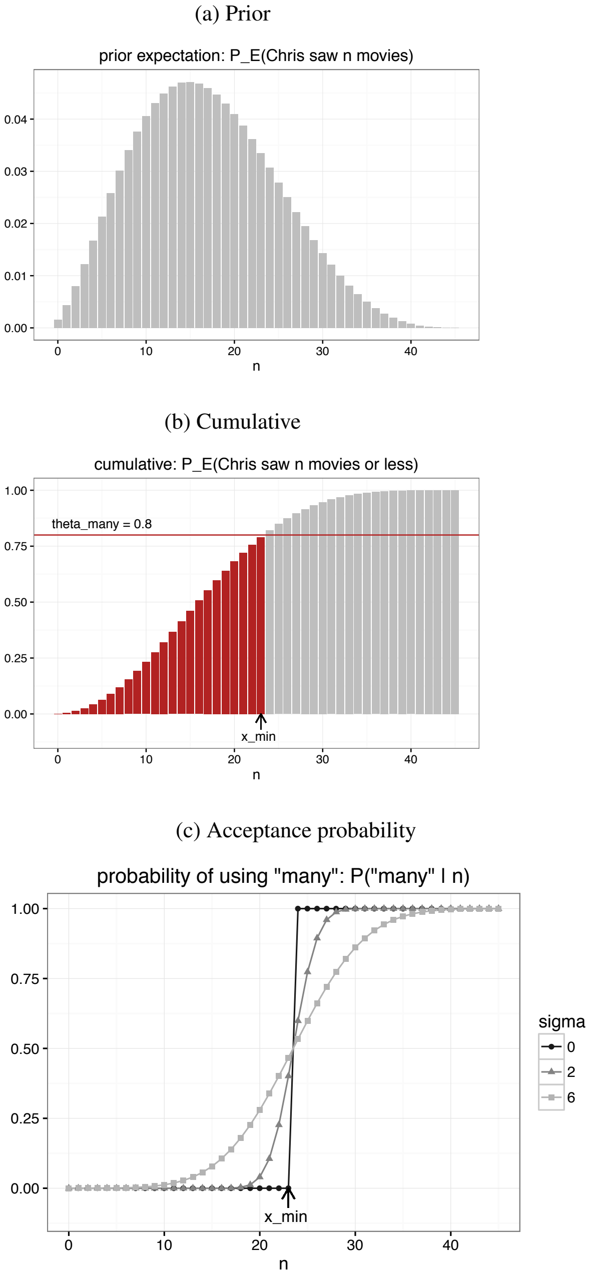
One concrete proposal of how $x_{\text{min}}$ and $x_{\text{max}}$ might be identified is presented by Fernando and Kamp (1996). The "cardinal surprise reading" of *few* and *many* in sentences like (1) is an intentional comparison between the actual number of movies that Chris saw last year and a probabilistic belief $P_E$ about the expected number of watched movies in some contextually provided *comparison class*. The *for*-phrase in (1) triggers a comparison class of US males. The prior expectation $P_E$ is highly context-dependent. In contrast, $\theta_{\text{few}}$ and $\theta_{\text{many}}$ are context-independent. They are fixed thresholds on the cumulative distribution of $P_E$. Truth conditions of the surprise reading of sentences like (1) are given in (5).

(5) a. $[\![\text{Few As are B}]\!] = 1$ iff $|A \cap B| \leq x_{\text{max}}$ where $x_{\text{max}} = \max\{n \in \mathbb{N} \mid P_E(|A \cap B| \leq n) < \theta_{\text{few}}\}$

b. $[\![\text{Many As are B}]\!] = 1$ iff $|A \cap B| \geq x_{\text{min}}$ where $x_{\text{min}} = \min\{n \in \mathbb{N} \mid P_E(|A \cap B| \leq n) > \theta_{\text{many}}\}$

From (5b), entities which have properties $A$ and $B$ can be described as "many" if their cardinality is at least $x_{\text{min}}$. $x_{\text{min}}$ is the lowest number for which the cumulative density mass of prior expectation $P_E$ about the number of As with property B is higher than the threshold $\theta_{\text{many}}$. In other words, "Many As are B" is a true description of cardinalities which are surprisingly high with respect to the contextually given $P_E$ and the context-independent threshold $\theta_{\text{many}}$ on $P_E$.

To illustrate, consider the example in Figure 1 for the *many*-sentence in (1). Prior expectations $P_E$ could look like in Figure 1a: they would assign a probability to any natural number $n$, indicating how likely we think it is that Chris saw $n$


(a) Prior

(b) Cumulative

(c) Acceptance probability

Figure 1: Illustration of a surprise-based semantics

movies last year. Figure 1b shows the cumulative distribution of the distribution in Figure 1a. If $\theta_{\text{many}}$ was fixed to, say, 0.8, then the semantics would identify $x_{\text{min}}$ to be 23. Accordingly, for this $P_E$, the *many*-sentence in (1) would be false for any $n < 23$ and true for any $n \geq 23$. Schöller and Franke (2015) present evidence for the fixed threshold hypothesis by identifying fixed values for $\theta_{\text{few}}$ and $\theta_{\text{many}}$, which correctly predict the applicability of *few* and *many* in different contexts, given experimentally measured prior expectations.

## Surprisingly: Intensifier or Marker of Surprise?

Two views are prima facie plausible for the meaning contribution of the adverb *surprisingly*. On the one hand, *surprisingly* can be taken to intensify the meaning of *few* and *many* just like other intensifiers like *incredibly* or *very* do. As a result, *surprisingly many* might be associated with a threshold $\theta_{\text{surpr. many}}$ higher than $\theta_{\text{many}}$. The contrasting view is to classify *surprisingly* as a marker of the surprise reading, which overtly marks that truth-conditions must draw on a threshold on a measure of surprise. This view is supported by the semantic literature which suggests that "being surprisingly tall comes to mean taller than expected" (Nouwen, 2011, 154).

Note that our hypotheses for *surprisingly* apply to sentences with a salient cardinal surprise reading and a restricted comparison class. To discriminate between the two views on *surprisingly*, we deduce two experimentally testable hypotheses. Another two auxiliary hypotheses are tested alongside to complement our understanding of modified *few* and *many*, see Table 1. In what follows, we spell out these general hypotheses in terms of their predictions about the threshold values $\theta_{\text{few}}$ and $\theta_{\text{many}}$ as assumed by Fernando and Kamp (1996) and test them with a computational model which infers these threshold values on the basis of experimental data.

**Salient surprise reading.** We cannot exclude that *few* and *many* may also denote a small or large cardinality, independent of prior expectations. To test the auxiliary assumption that the most salient readings of our experimental test sentences (see Appendix) are cardinal surprise readings given the comparison class for which we measure subjects' prior expectations (see below), we contrast sentences with bare *few* and *many* with sentences modified by the *compared to* phrase in (2) which makes the relevant expectations overt. It is necessary to test this because if *few* and *many* did not have the intended surprise reading, differences between *few/many* and *surprisingly few/many* could be due to different readings and possibly different threshold values associated with them. Alongside *few* and *many*'s intrinsic surprise reading, we test another related assumption: the *for*- phrase used to mark the comparison class triggers the same prior expectations $P_E$ as the *compared to* phrase which openly addresses expectations, see (7).

**Marker of surprise.** If the function of *surprisingly* is to mark a cardinal surprise reading, thresholds are the same as for unmodified *few/many*, where these cardinal surprise readings are most salient anyway (see above). Furthermore, sentences with *surprisingly* should not be different from sentences with *compared to*, as in (2).

**Intensifier.** Modification by *surprisingly* raises the threshold of *many* and makes it applicable to a smaller range of cardinalities, resulting in a stronger statement than the alternative with bare *many*. *Few*'s threshold decreases.

**Bennett & Goodman.** The intensifier hypothesis is in line with work by Bennett and Goodman (2015) who explain the

|  | hypothesis |  |  |
|---|---|---|---|
|  | **intensifier** | **marker of surprise** | **salient surprise reading** |
| predictions | *many* $\leq$ *surprisingly many*<br>*few* $\geq$ *surprisingly few*<br>*surprisingly* = *incredibly* | *many* = *surprisingly many*<br>*few* = *surprisingly few*<br>*surprisingly* = *compared to* | *many* = *compared to... many*<br>*few* = *compared to... few* |
| results | *few*: × *many*: ✓ | *few*: ✓ *many*: × | *few*: ✓ *many*: ✓ |

Table 1: Hypotheses and results

strength of an intensifying degree adverb as "pragmatic inference based on differing cost [(their length and frequency)] rather than differing semantics" (p. 1). However, they do not test *surprisingly*. From the adverbs tested by Bennett and Goodman (2015), *incredibly* comes closest to *surprisingly*, as they have the same number of syllables and the most similar frequency in an updated version of the corpus Bennett and Goodman (2015) used, the Google Web 1 T 5grams corpus (4,987,059 occurrences as compared to 4,373,670 occurrences of *surprisingly*). Following Bennett and Goodman (2015), we hypothesize that the thresholds of *surprisingly few/many* are roughly the same as for *incredibly few/many*.

## Experiments

To test the hypotheses in Table 1, two experiments were conducted to gather acceptability ratings of sentences with (modified) *few* and *many* and to measure participants' prior expectations. Prior expectations will be input to the computational model, which is presented in the next section.

### Experiment 1: Prior elicitation

**Design.** To get an empirical estimate of participants' prior expectations, we used a *binned histogram task*. Participants saw descriptions of a context as in (6a) and a question as in (6b). Subjects were presented with 15 intervals, whose ranges were determined by a pre-test (in which we asked for the most likely, lowest and highest possible cardinality). Subjects rated the likelihood that the true value lies in each of the intervals, by adjusting a slider labeled from "extremely unlikely" to "extremely likely." For example, they would adjust a slider each for the probability that Chris saw 0–2, 3–5, ..., 39–41 or more than 42 movies last year.

(6) **Prior elicitation example**

a. BACKGROUND: Chris is a man from the US.

b. How many movies do you think he saw last year?

**Participants.** 80 subjects were recruited via Amazon's Mechanical Turk with US-IP addresses.

**Materials & Procedure.** After reading instructions, each subject saw all of the 14 experimental items (see Appendix), one after another. For each item, the 15 intervals were presented horizontally on the screen and paired with a vertical slider. Participants had to adjust or at least click on each slider before being able to proceed.

**Results.** Two participants were excluded for not being native speakers of English. For each item, each participant's ratings were normalized and these normalized ratings were then averaged across participants. We understand these probability distributions $P_E$, see Figure 2, as approximations of the beliefs held in the population (Franke et al., 2016).

### Experiment 2: Judgment task

**Design.** In a binary judgment task we measured acceptance of sentences with *few* and *many* with and without modifiers (*surprisingly, incredibly* or *compared to*). Participants were presented with a context which introduced a situation and an interval as in (7a). The interval was randomly chosen from 8 of the 15 intervals from the prior elicitation task (see Appendix). We presented only four low intervals for *few* and four high intervals for *many* to avoid a large number of combinations. The context was described by a statement as in (7b) which contained either *few* or *many*. We elicited data of four groups of participants which each saw a different modifier.

(7) **Production study example**

a. CONTEXT: Chris is a man from the US who saw [0–2 | 6–8 | ... | 42 or more] movies last year.

b. STATEMENT: [For | Compared to what you would expect for] a man from the US, Chris saw [- | surprisingly | incredibly ] [few | many] movies last year.

**Materials & Procedure.** Each participant was randomly assigned to one modifier condition (unmodified, *compared to* construction, *surprisingly*, *incredibly*). After reading a short explanation of the task, each subject saw all of the 14 contexts from the Appendix one after another in random order. Sentences with unmodified *few* and *many* or *incredibly* or *surprisingly* were introduced by a *for*-phrase which made the intended comparison class overt. The fourth group saw a *compared to* phrase which additionally made expectations salient. For each context, a quantity word and one of its four associated intervals were assigned randomly. Participants had to click on one of two radio buttons labeled with TRUE or FALSE before being able to proceed to the next item.

**Participants.** We recruited 787 participants with US-IP addresses via Amazon's Mechanical Turk, among them 301 participants in the unmodified condition and 162 participants

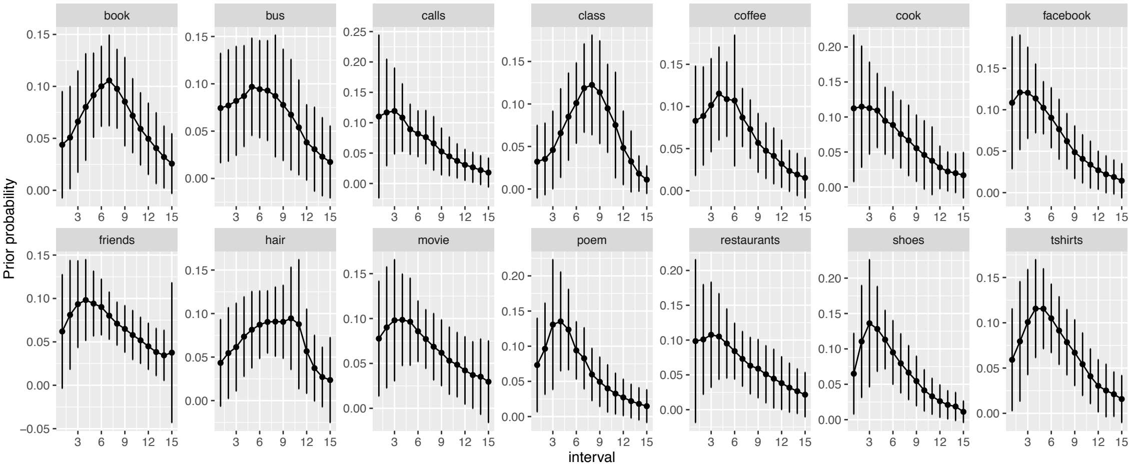


Figure 2: Empirically measured prior expectations. Error bars are estimated 95% confidence intervals.

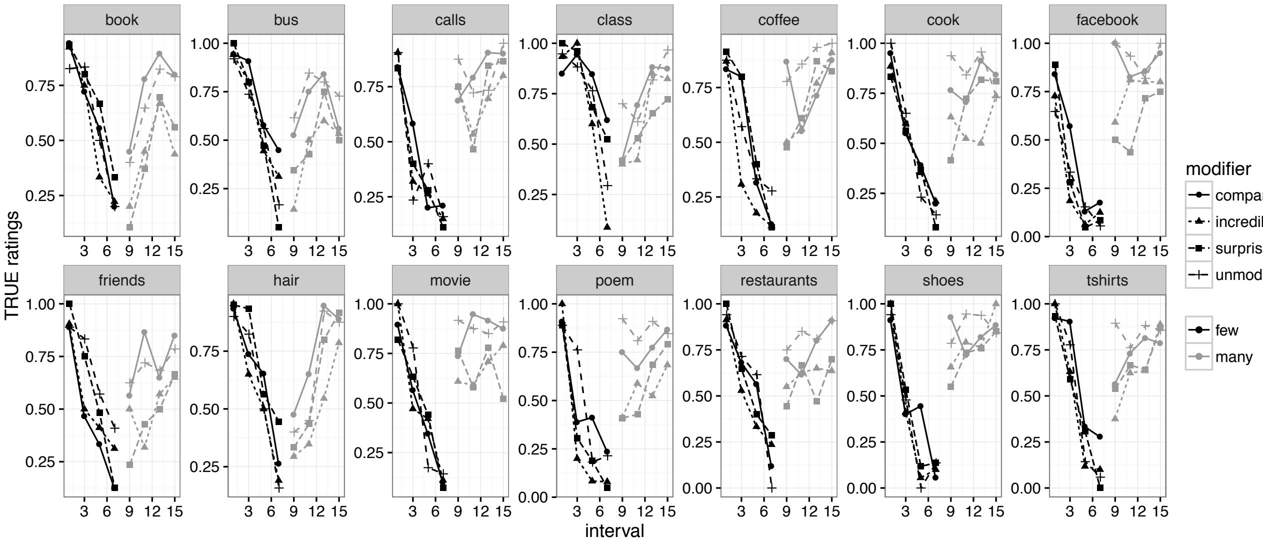


Figure 3: Proportion of TRUE answers from Experiment 2.

each in the other three conditions. The unmodified condition had more participants because it was part of a previous experiment in which we presented 8 of 15 intervals for both *few* and *many*. For the analysis only data from those intervals presented in the other three conditions was used.

**Results.** Data was excluded of 25 participants who reported not to be native speakers of English or to not having understood the task. Figure 3 shows the proportion of TRUE answers.

For each of the quantity words *few* and *many* we specified a linear mixed effects regression model predicting the proportional acceptance of statements as in (7b). During a guided search through the model space, we started out with a model containing only the random effect ITEM and added fixed effects if this significantly increased the model's fit to the data (measured by AIC).

For *many*, the final model includes the fixed effects INTERVAL and MODIFIER and their interaction. Significantly more participants accepted the statements for higher intervals ($\beta = 0.02, SE = 0.007, p < 0.01$). The modification of *many* by *surprisingly* leads to a lower acceptance ($\beta = -0.59, SE = 0.12, p < 0.001$) than of sentences with unmodified *many*. This suggests that *surprisingly* intensifies the meaning of *many*. The same is the case for sentences with *incredibly*, which were also rated lower than unmodified *many* ($\beta = -0.53, SE = 0.12, p < 0.001$). There is no difference between sentences with a *compared to* phrase and unmodified *many* ($\beta = -0.17, SE = 0.12, p < 0.15$), which suggests that *many* receives a surprise reading in both cases. *Surprisingly* and *compared to* are rated significantly different

($\beta = -0.42, SE = 0.12, p < 0.001$), but there is no difference between *surprisingly* and *incredibly*. Furthermore, there is a significant interaction between INTERVAL and MODIFIER for *surprisingly* ($\beta = 0.03, SE = 0.01, p < 0.001$) and *incredibly* ($\beta = 0.02, SE = 0.01, p < 0.01$).

For *few*, the final model, obtained by the same procedure, includes the fixed effects INTERVAL and MODIFIER. The proportion of participants accepting the statement is significantly lower for higher numbers ($\beta = -0.12, SE = 0.004, p < 0.001$). Among the modifiers only *incredibly* is significantly different from bare *few* ($\beta = -0.05, SE = 0.02, p < 0.05$); for *surprisingly* and *compared to* this is not the case. No significant difference between *surprisingly* and *compared to* is found, but *incredibly* is rated significantly lower than *surprisingly* ($\beta = -0.05, SE = 0.02, p < 0.05$).

These results are expected under the "salient surprise reading" hypothesis. While *surprisingly* seems to behave like an intensifier for *many*, for *few* it seems to redundantly mark surprise.

## Computational Model

The regression models reported above include a random effect for items but do not constrain these to reflect prior expectations. Moreover, regression models do not predict judgments as a function of thresholds on expectations. It is therefore insightful to complement regression modeling with an explicit theory-driven model of a possible data-generating process. We use the computational model of Schöller and Franke (2015) for this purpose. The model takes empirically measured prior expectations as input and treats $\theta_{\text{[i]few}}$ and $\theta_{\text{[i]many}}$ for each modifier condition $i$ (unmodified, *surprisingly, incredibly, compared to*) as latent parameters, whose values will be estimated to fit experimental data. The model specifies a likelihood function $P(\text{Observation} \mid \theta_{\text{[i]many}}, \theta_{\text{[i]few}})$ which assigns to values of latent parameters a probability of seeing a particular experimental observation. Bayesian inference is one way to infer plausible threshold values, given the likelihood function and a prior:

$$P(\theta_{\text{[i]many}}, \theta_{\text{[i]few}} \mid \text{O}) \propto P(\theta_{\text{[i]many}}, \theta_{\text{[i]few}}) P(\text{O} \mid \theta_{\text{[i]many}}, \theta_{\text{[i]few}})$$

Our goal, then, is to see for each modifier which pairs of threshold values $\theta_{\text{[i]many}}$ and $\theta_{\text{[i]few}}$ are likely given the data. We estimate the a posteriori credible threshold values and compare how similar they are across conditions. We focus on *many* in the exposition, but the case for *few* is parallel. Straightforwardly, (5) translates into a probabilistic rule $P(\text{“[modifier } i\text{] many”} \mid n, P_E \; ; \; \theta_{\text{[i]many}}) = \delta_{n \geq x_{\text{min,i}}}$, where $x_{\text{min,i}}$ is derived from $P_E$, as in (5), based on $\theta_{\text{[i]many}}$. This is a degenerate probabilistic rule because it maps the applicability of "many" to 0 and 1 only. To allow for noise, we look at a parameterized, smoothed-out version.

$$P(\text{“}[i]\text{ many”} \mid n, P_E; \theta_{\text{[i]many}}, \sigma_j) = \sum_{k=0}^{n} \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} \mathcal{N}(y; x_{\text{min,i}}, \sigma_j) \mathrm{d}y$$

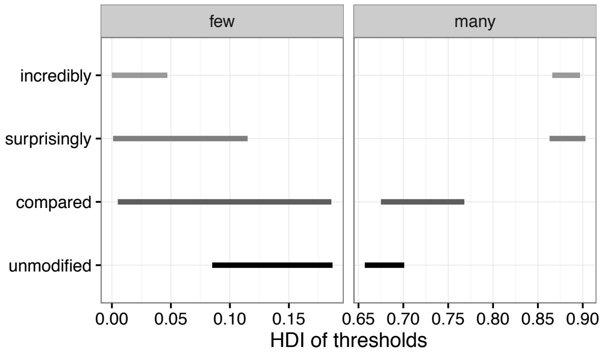


Figure 4: Estimated 95% credible intervals for $\theta_{\text{few,i}}$ & $\theta_{\text{many,i}}$

Here, $\sigma_j$ is another free model parameter that regulates the steepness of the curve, and $\mathcal{N}(y; x_{\text{min,i}}, \sigma_j)$ is the probability density of $y$ under a normal distribution with mean $x_{\text{min,i}}$ and standard deviation $\sigma_j$. This rule predicts noisy acceptability ratings under a surprise-based semantics where the amount of noise is controlled by $\sigma_j$, see Figure 1c. Noise can be caused by uncertainty about the exact shape of $P_E$ and the amount of uncertainty differs across contexts. This is why we allow an individual value of $\sigma_j$ for each context $j$. Furthermore, we assume that the parameter values $\theta_{\text{[i]many}}$, $\theta_{\text{[i]few}}$ and $\sigma_j$ are independent of each other and that they have uniform priors over an interval that is large enough to accommodate a range of plausible values without weighting them.

$$P(\theta_{\text{[i]many}}, \theta_{\text{[i]few}}, \sigma_j) = \text{Uniform}_{[0;1]}(\theta_{\text{[i]many}}) \cdot \\ \text{Uniform}_{[0;1]}(\theta_{\text{[i]few}}) \cdot \text{Uniform}_{[0;10]}(\sigma_j)$$

To approximate the joint posterior distribution, we used MCMC sampling, as implemented in JAGS (Plummer, 2003). We collected 10,000 samples from 2 MCMC chains after a burn-in of 10,000. This ensured convergence, as measured by $\hat{R}$ (Gelman & Rubin, 1992). Figure 4 shows the estimated 95% credible intervals for the marginalized posteriors over thresholds per modifier. Where intervals (clearly) do not overlap, there is reason to believe that thresholds differ. For example, $\theta_{\text{surpr.many}} \in [0.863, 0.903]$ tells us that *surprisingly many* describes cardinalities which are higher than at least 86% of the cumulative density mass of $P_E$. This threshold is higher than bare *many*'s, $\theta_{\text{many}} \in [0.657, 0.701]$. Taken together, the model predicts that *surprisingly many* is restricted to describe higher cardinalities than unmodified *many*.

## Discussion and Conclusions

Table 1 summarizes the results from regression and theory-driven modeling. The data supports the "salient surprise reading" hypothesis assumed by Fernando and Kamp (1996) and suggests that an expectation-based reading is the canonical interpretation of cardinal *few* and *many* in our test sentences. There is no difference between unmodified sentences and sentences in which expectations are made salient by a *compared to* phrase.

For *surprisingly*, the picture is less clear. Sentences with *many* provide support for the "intensifier" hypothesis. Speakers prefer it for higher cardinalities than those which render unmodified *many* or sentences with a *compared to* construction true. Furthermore, we do not find a difference with *incredibly*. When combined with *few*, however, *surprisingly* does not appear to be an intensifier. Sentences with *few*, *surprisingly few* and *compared to* are rated equally, speaking in favor of a "marker of surprise" hypothesis. For the comparison between *surprisingly* and *incredibly*, we get conflicting results from the regression and the theory-driven model. The regression analysis finds that *incredibly few* is rated lower than *surprisingly few*, but the computational model identifies an overlap in the estimated credible intervals. However, we want to once more stress that we are here comparing conclusions based on models which are decidedly different. Whereas the computational model is theory-driven and includes experimentally measured prior expectations, the regression model only looks at numerical differences in the ratings. Ultimately, we believe in the computational model.

Keeping in mind that *few* only applies to small cardinalities, the lack of a difference could also be due to a floor effect. This is where future research should tie in. *Few* should be presented in contexts like **book** or **facebook**, in which large cardinalities are plausible and *few* can operate away from 0. Additionally, the presented intervals should be more fine-grained. A follow-up study as well as further discussion of the semantic differences between *few* and *many* are presented in Schöller (2017).

## Acknowledgments

MF and AS's work was supported by the Priority Program XPrag.de (DFG Schwerpunktprogramm 1727). MF's work was supported by the Institutional Strategy of the University of Tübingen (DFG, ZUK 63).

## Experimental material

1. **book** — A friend's favorite book has been published only recently (and has [0-40, 81-120, 161-200, 241-280, 321-360, 401-440,481-520, 560 or more] pages).
2. **bus** — Vehicle No. 102 is a school bus (which has seats for [0-4, 10-14, 20-24, 30-34, 40-44, 50-54, 60-64, 70 or more] passengers).
3. **calls** — Lisa is a woman from the US (who made [0-4, 10-14, 20-24, 30-34, 40-44, 50-54, 60-64, 70 or more] phone calls last week).
4. **class** — Erin is a first grade student in primary school. (There are [0-2, 6-8, 12-14, 18-20, 24-26, 30-32, 36-38, 42 or more] children in Erin's class.)
5. **coffee** — Andy is man from the US (who drank [0-1, 4-5, 8-9, 12-13, 16-17, 20-21, 24-25, 28 or more] cups of coffee last week).
6. **cook** — Tony is a man from the US (who cooked himself [0-3, 8-11, 16-19, 24-27, 32-35, 40-43, 48-51, 56 or more] meals at home last month).
7. **facebook** — Judith is a woman from the US (who has [0-69, 140-209, 280-349, 420-489, 560-629, 700-769, 840-909, 980 or more] Facebook friends).
8. **friends** — Lelia is a woman from the US (who has [0-1, 4-5, 8-9, 12-13, 16-17, 20-21, 24-25, 28 or more] friends).
9. **hair** — Betty is a woman from the US (who washed her hair [0-2, 6-8, 12-14, 18-20, 24-26, 30-32, 36-38, 42 or more] times last month).
10. **movie** — Chris is a man from the US (who saw [0-2, 6-8, 12-14, 18-20, 24-26, 30-32, 36-38, 42 or more] movies last year).
11. **poem** — A friend wants to read you her favorite poem (which has [0-3, 8-11, 16-19, 24-27, 32-35, 40-43, 48-51, 56 or more] lines).
12. **restaurants** — Sarah is a woman from the US (who went to [0-3, 8-11, 16-19, 24-27, 32-35, 40-43, 48-51, 56 or more] restaurants last year).
13. **shoes** — Melanie is a woman from the US (who owns [0-2, 6-8, 12-14, 18-20, 24-26, 30-32, 36-38, 42 or more] pairs of shoes). — intervals:
14. **tshirts** — Liam is a man from the US (who has [0-2, 6-8, 12-14, 18-20, 24-26, 30-32, 36-38, 42 or more] T-shirts).

## References

Bennett, E., & Goodman, N. D. (2015). Extremely costly intensifiers are stronger than quite costly ones. In *Proceedings of CogSci* (pp. 226–231).

Clark, H. H. (1991). Words, the world, and their possibilities. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The perception of structure: Essays in honor of Wendell R. Garner* (pp. 263–277). American Psychological Association.

Fernando, T., & Kamp, H. (1996). Expecting many. In T. Galloway & J. Spence (Eds.), *Linguistic society of america SALT* (pp. 53–68). Ithaca, NY: Cornell University.

Franke, M., Dablander, F., Schöller, A., Bennett, E. D., Degen, J., Tessler, M. H., ... Goodman, N. D. (2016). What does the crowd believe? A hierarchical approach to estimating subjective beliefs from empirical data. In *Proceedings of CogSci* (pp. 2669–2674).

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 457–472.

Nouwen, R. (2011). Degree modifiers and monotonicity. In *Vagueness and language use* (pp. 146–164). Springer.

Partee, B. (1989). Many quantifiers. In J. Powers & K. de Jong (Eds.), *5th eastern states conference on linguistics (escol)* (pp. 383–402).

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In K. Hornik, F. Leisch, & A. Zeileis (Eds.), *Proceedings of the 3rd international workshop on distributed statistical computing.*

Sanford, A. J., Moxey, L. M., & Paterson, K. (1994). Psychological studies of quantifiers. *Journal of Semantics*, *11*(3), 153–170. doi: 10.1093/jos/11.3.153

Schöller, A. (2017). *How many are many? Exploring context-dependence with probabilistic computational models.* (Unpublished doctoral dissertation)

Schöller, A., & Franke, M. (2015). Semantic values as latent parameters: Surprising few & many. In S. D'Antonio, M. Moroney, & C. R. Little (Eds.), *Proceedings of SALT* (Vol. 25, pp. 143–162). doi: 10.3765/salt.v25i0.3058