Semantic values as latent pragmatic parameters: surprising *few* and *many*

Background. The context-independent meaning component of vague quantifiers *few* and *many* is highly elusive, especially under their "cardinal surprise readings" (c.f. Partee, 1988; Fernando and Kamp, 1996) exemplified in (1).

(1) Joseph eats few / many burgers. $\rightsquigarrow$ Joseph eats less / more burgers than expected.

An intuitive semantics for (1) was first suggested tentatively by Clark (1991), floated in much subsequent psychological work (e.g. Moxey and Sanford, 1993), and formally spelled out by Fernando and Kamp (1996). According to the Clark-Fernando-Kamp (CFK) semantics, as we will call it here, the target reading of *few* and *many* in (1) is intensional and compares the actual number of burgers that Joseph eats to a probabilistic belief P about the expected number of consumed burgers in the relevant comparison class. While the prior expectation P is highly context-dependent, the context-independent lexical meaning contribution of *few* and *many* is a fixed threshold on the cumulative distribution of P, similar to degree semantics for gradable adjectives (c.f. Solt, 2011):

(2) $P(\{w \mid \text{Joe eats less burgers in } w \text{ than in the actual world}\}) \quad < \theta_{\text{few}} \quad / \quad > \theta_{\text{many}}$

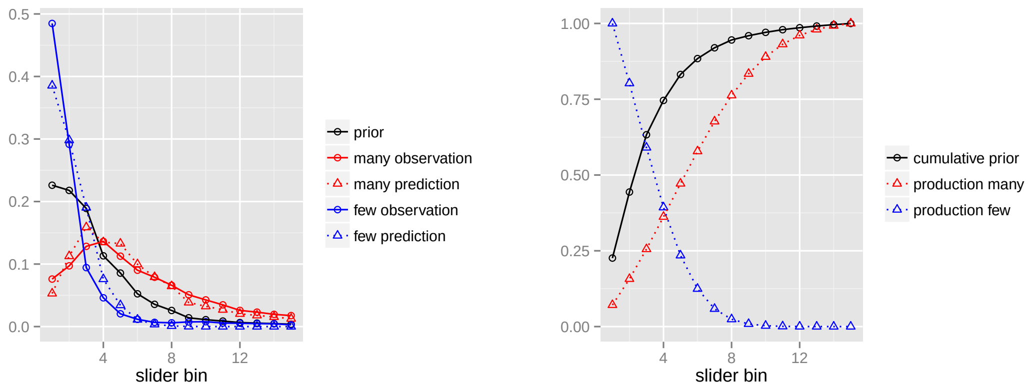
Despite the clear intuitive appeal, the CFK semantics has a severe drawback: it is almost impossible to verify by introspection, and hard to test experimentally. This is particularly disappointing because the case does promise to give important insights into the interaction between pragmatically variable meaning elements like P and fixed semantic values like θ_{few} and θ_{many}. We therefore apply recent experimental methodology to elicit subjects' prior expectations over arbitrary real world events (e.g. Kao et al., 2014) in order to test the CFK semantics of *few* and *many*. By doing so, we present a case for thinking of semantic values, in particular those that are inaccessible to even trained introspection, as determined by their explanatory success within a pragmatic model that aims to predict experimental data — in slogan form: semantic values as latent pragmatic parameters.

Methods. 60 participants from the USA were recruited via Amazon's Mechanical Turk and reimbursed with 0.60$. 15 items were presented in three conditions (prior/few/many) in a random order. Every participant saw each item only in one condition. An example:

(3) (a) Joe is a man from the US. **prior**
(b) Joe is a man from the US who eats few burgers. **few**
(c) Joe is a man from the US who eats many burgers. **many**
How many burgers do you think Joe eats per month?

Participants were presented with 15 slider-interval pairs (the size of the bins depended on the respective item, determined by a pre-test) and had to rate how likely they thought the speaker referred to a number in each interval. Items without a quantifier were included as a prior elicitation task. The 15 ratings per item, indicating the likelihood of each interval, were normalized by subject-item-condition and subsequently averaged over item-condition. The data of 9 participants had to be excluded because they were not native speakers of English or because their ratings were incomplete or obviously uncooperative.

Model. Taking the results from the prior condition as input, our goal is to estimate the single best value for θ_{few} / θ_{many} that best explains the data from the *few* / *many* condition across all items. Since the latter are interpretation conditions, we model their data as the outcome of a Bayesian inference to the most likely world state (slider bin) based on the assumption that the speaker's utterance ("few" or "many") was true under a CFK semantics. This gives us a straightforward prediction function from Bayes' rule: the probability that the listener assigns to the i^{th} slider bin given utterance $u \in \{\text{“few”}, \text{“many”}\}$



(a) Data and predictions for the "burgers" item under best-fit parameters.

(b) Corresponding speaker production rule, with cumulative prior.

for experimental item j is $P(\text{bin}_i \mid u, \text{item}_j; \sigma_j) \propto P(\text{bin}_i; \text{item}_j) \cdot P(u \mid \text{bin}_i, \text{item}_j; \sigma_j)$, where $P(\text{bin}_i; \text{item}_j)$ is the empirically measured prior and $P(u \mid \text{bin}_i, \text{item}_j; \sigma_j)$ is the production probability under a CFK semantics with item-dependent Gaussian noise with standard deviation σ_j around the semantic threshold:

$$P(\text{“many”} \mid \text{bin}_i, \text{item}_j; \sigma_j) = \sum_{k \leq i} \mathcal{N}^{\mathbb{N}}_{[1;15]}(k; \mu_j, \sigma_j)\,,$$

$$\text{with } \mu_j = \arg \min_{k \in [1;15]} \sum_{l \leq k} P(\text{bin}_k; \text{item}_j) > \theta_{\text{many}}\,,$$

where $\mathcal{N}^{\mathbb{N}}_{[1;15]}$ is a discretized, truncated normal distribution. (The formulation for *few* is parallel, in reversed direction; see Figure (a) for examples.) This production rule implements a smoothed-out step function around the semantic threshold, with slack parameter σ_j modeling vagueness (e.g., due to uncertainty about the precise prior; see Figure (b)). We determined the "best" values for θ_{few} and θ_{many} as those that, together with a vector of σ_j's, minimized the Kullback-Leibler divergence of the prediction $P(\text{bin}_i \mid u, \text{item}_j; \sigma_j)$ to the empirically measured distributions from the "few" and "many" conditions. The resulting best estimates $\theta_{\text{few}} \approx .371$ and $\theta_{\text{many}} \approx .635$ yield a good fit to the data (correlation of observation and prediction at all data points: $r \approx .87$, $p < 0.01$; RMSD $\approx .036$; mean KL-divergence over all items and conditions $\approx .13$).

Discussion. Our aim is not to defend the CFK semantics against competitors, or to claim that our probabilistic modeling approach is superior to any alternative. Rather this exercise in probabilistic modeling makes more general, conceptually relevant contributions. First, maintaining a context-independent semantic threshold for the applicability of vague quantifiers *few* and *many* is possible, if we allow for contextual variation in prior expectations. Second, in line with like-minded recent contributions, said prior expectations can be measured and utilized for linguistic theorizing. Third, it may serve as a concrete example for the foundational point that semantic values can be conceived as latent explanatory parameters in serious-enough pragmatic models of experimental data.

References. Clark: *Words, the World, and their Possibilities* (1991). Fernando & Kamp: *Expecting Many* (1996). Kao et al.: Nonliteral understanding of number words (2014). Moxey & Sanford: *Communicating Quantities* (1993). Partee: *Many Quantifiers* (1988). Solt: *Vagueness in Quantity* (2011).