**Lassiter & Goodman (2013)**

"Bayesian modeling makes it possible to combine logical and probabilistic reasoning seamlessly, opening up new opportunities for exploring how listeners use context and background knowledge to construct rich context-sensitive interpretations in the presence of uncertainty." (p. 605)

| Ambiguity | Vagueness |
|---|---|
| Ambiguous terms have two or more *distinct* meanings. | Vague terms give rise to:<br>• Sorites paradox<br>• Borderline cases |

Lassiter & Goodman (2013) pursue 2 main goals:

1) Extend Frank & Goodman's (2012) *Rational Speech Act (RSA)* model to incorporate degree semantics
2) Explain phenomena associated with the relative/absolute distinction

Extending the *RSA* to incorporate degree semantics

| | Original *RSA* model | Extended *RSA* model |
|---|---|---|
| $L_0$ | $P_{L_0}(A \mid u) = P_{L_0}(A \mid [\![u]\!] = 1)$<br>(10) in L&G (2013) | $P_{L_0}(A \mid u, V) = P_{L_0}(A \mid [\![u]\!]^V = 1)$<br>(11) in L&G (2013) |
| $S_1$ | $U(w; r_S, C) = I(w; r_S, C) - D(w)$<br>(S2) in F&G (2012)<br>$P(w \mid r_S, C) = \frac{e^{-(-\log(\lvert w \rvert^{-1}))}}{\sum_{w' \in V \text{ st. } w'(r_s)=\text{true}} e^{-(-\log(\lvert w' \rvert^{-1}))}},$<br>(S4) in F&G (2012) | $\mathbb{U}_{S_1}(u; A, V) = \log(P_{L_0}(A \mid u, V)) - C(u)$<br>(12) in L&G (2013)<br>$P_{S_1}(u \mid A, V) \propto \exp(\alpha \times \mathbb{U}_{S_1}(u; A, V))$<br>(13) in L&G (2013) |
| $L_1$ | $P(r_s \mid w, C) = \frac{P(w \mid r_s, C) P(r_s)}{\sum_{r' \in C} P(w \mid r', C) P(r')}$<br>(1) in F&G (2012) | $P_{L_1}(A, V \mid u) \propto P_{S_1}(u \mid A, V) \times P_{L_1}(A, V)$<br>$\propto P_{S_1}(u \mid A, V) \times P_{L_1}(A) \times P_{L_1}(V)$<br>$P_{L_1}(A, V \mid u) \propto P_{S_1}(u \mid A, V) \times P_{L_1}(A)$<br>(14 and 15) in L&G (2013) |

The degree semantics L&G assume

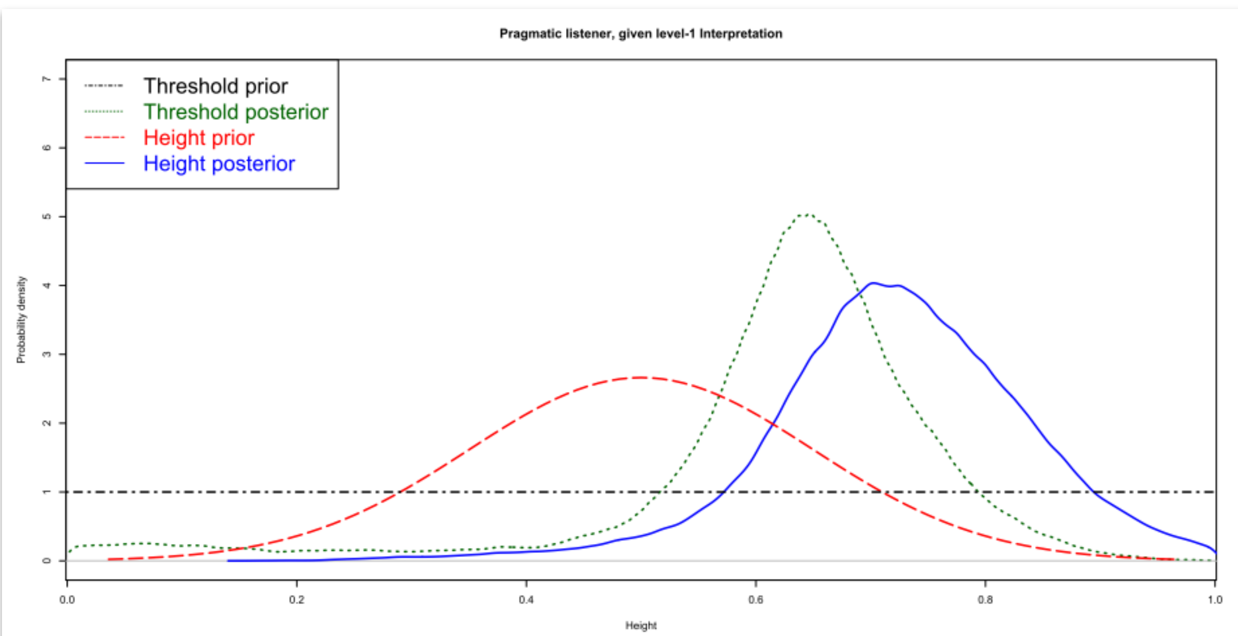
| | |
|---|---|
| Lexical entry (of adjectives):<br>• Scale (e.g. 0-∞)<br>• Polarity | The meaning of an adjective is a Boolean function with **3 arguments**:<br>• *x: an entity;*<br>• $\mu_A(x)$: *x's place on A's scale;*<br>• $\theta_A$: *the threshold parameter*<br><br>**NOTE:** Here, *A* is short for *adjective*; in the model above, L&G use *A* to refer to Al's actual height, i.e. $\mu_{tall}(Al)$ in the present notation |
| | |

| (3) $[\![A]\!] = \lambda\theta_A \lambda x[\mu_A(x) > \theta_A]$ | |
|---|---|
| **Silent-morpheme account** | **Type-shifter account** |
| (4) a. $[\![tall]\!]^{\mathbf{s}_{tall}, \mathbf{s}_{big}, \mathbf{s}_{heavy}, \dots} = \lambda\theta_{tall}\lambda x[\mu_{tall}(x) > \theta_{tall}]$<br>b. $[\![POS]\!]^{\mathbf{s}_{tall}, \mathbf{s}_{big}, \mathbf{s}_{heavy}, \dots} = \lambda A \lambda x[A(\mathbf{s}_A)(x)]$<br>c. $[\![POS\ tall]\!]^{\mathbf{s}_{tall}, \mathbf{s}_{big}, \mathbf{s}_{heavy}, \dots} = \lambda x[\mu_{tall}(x) > \mathbf{s}_{tall}]$<br>d. $[\![Al\ is\ POS\ tall]\!]^{\mathbf{s}_{tall}, \mathbf{s}_{big}, \mathbf{s}_{heavy}, \dots} = \mu_{tall}(\mathbf{Al}) > \mathbf{s}_{tall}$ | (5) a. $[\![tall]\!] = \lambda\theta_{tall}\lambda x[\mu_{tall}(x) > \theta_{tall}]$<br>b. $[\![POS]\!] = \lambda A \lambda x \lambda\theta_A[A(\theta_A)(x)]$<br>c. $[\![POS\ tall]\!] = \lambda x \lambda\theta_{tall}[\mu_{tall}(x) > \theta_{tall}]$<br>d. $[\![Al\ is\ POS\ tall]\!] = \lambda\theta_{tall}[\mu_{tall}(\mathbf{Al}) > \theta_{tall}]$ |

- *Al is tall.*
  - $x = Al$
  - $\mu_{tall}(x) > \theta_{tall}$
  - $\theta_{tall}$ = indeterminate
  - ➔ The meaning of *tall* is not determined and so neither is that of *Al is tall*
  - ➔ The sentence *Al is tall* is neither true nor false, unless $\theta_{tall}$ is (contextually) fixed
- *Al is 2' tall.*
  - $x = Al$
  - $\mu_{tall}(x) > \theta_{tall}$
  - $\theta_{tall}$ = 2 ft
  - ➔ The meaning of *tall* is fixed semantically: taller or equal to 2 ft
  - ➔ The sentence *Al is 2' tall* is true iff *Al* is 2' or taller.

<u>Model predictions</u>

$P_{L1}(A, V | u = Al\ is\ tall)$:




Emergent properties:

- <u>Invariant core meanings</u>: quantitative effects of prior over *A* (and absence of qualitative effects) account for common denominator of *tall* in *tall man* and *tall house*

- Residual vagueness: Inferred knowledge/belief about Al's height (*A*) and the meaning of *tall* ($\theta_{tall}$) remains probabilistic; rationality produces categorical behavior given vague understanding
- Borderline cases: the applicability of *tall* to *Al* depends on the distance of Al's height (*A*) to the threshold ($\theta_{tall}$); maximal uncertainty arises when the modes of the inferred distributions for *A* and $\theta_{tall}$ are the same

The absolute/relative adjective distinction

| Absolute Adjectives (e.g. *dangerous/safe*) | Relative Adjectives (e.g. *tall/short*) |
|---|---|
| • Less context-sensitive | • Context-sensitive |
| • No borderline cases | • Give rise to borderline cases |
| • Less susceptible to Sorites paradox | • Susceptable to Sorites paradox |
| • Tend to live on bounded scales | • Tend to live on unbounded scales |

Main observation: Relative adjectives tend to have scales with no endpoints (unbounded), while absolute adjectives seem to require endpoints (bounded).

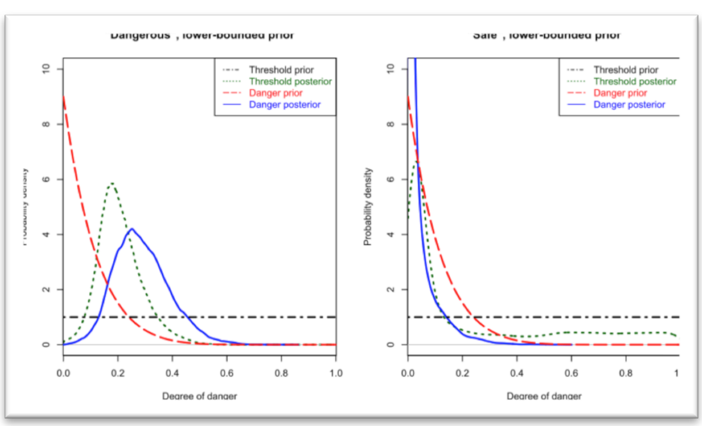
Kennedy's (2007) proposal: Scale boundedness determines interpretive differences between absolute and relative adjectives because the meanings of absolute adjectives are anchored to one of the endpoints of their scale.

Problems: exceptions to the above-listed tendencies (borderline cases; boundedness of vague adjective *bald*; one-sided boundedness of *cheap/expensive*, *short/tall*, etc.); remaining explananda (the role of conventions in the interpretation of bounded adjectives, e.g. *full wine glasses* vs. *full beer glasses*);

F&G's proposal: instead of scale *structure*, their model invokes the shape of priors as the source of the observed interpretive differences between relative and absolute adjectives

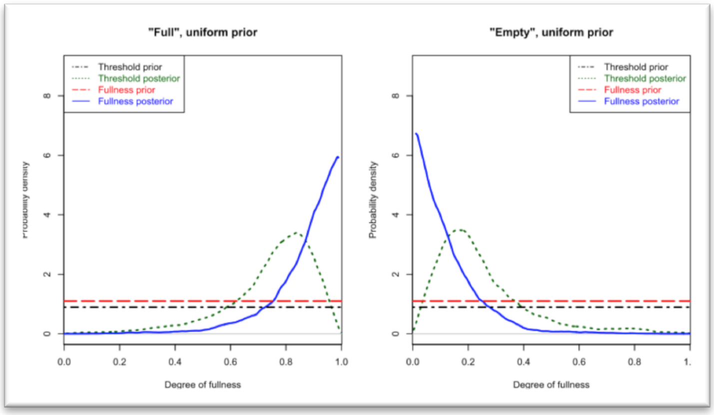
- Prototypical relative interpretations arise from priors with little probability mass at scale endpoints
- Prototypical absolute interpretations arise from priors with large portion of probability mass at scale endpoints

Min/max pairs (e.g. *dangerous/safe*) arise from Beta(1,9) prior:




(Figure 2 in L&G)

Max/max pairs (e.g. *empty/full*) arise from uniform prior:




(Figure 3 in L&G)

Emergent properties:

- There is *some* interpretive uncertainty about the applicability of *all* adjectives, but the amount and distribution of uncertainty depends on priors over adjective scale. This corresponds to the intuition that absolute adjectives are less vague than relative adjectives.
- There meaning of all adjectives is *somewhat* uncertain, but the amount of this uncertainty (i.e., variance in inferred threshold distribution) depends on priors.
- The applicability of adjectives (e.g. *empty/full*) varies with the context (e.g. beer vs. wine) because so does the threshold prior.

The (unsolved?!) Sorites Paradox

| The classical paradox | L&G's representation |
|---|---|
| ($P_1$) André the Giant is tall. | ($P_1$) *P*(*X is tall*) is high. |
| ($P_2$) Anyone who is 1mm shorter than someone who is tall, is tall. | ($P_2$) For all Y, *P*($Y_{m-1}$ *is tall* is true \| $u = Y_{m-1}$ *is tall*) Anyone who is 1mm shorter than someone who is tall, is tall. |
| (*C*) Danny DeVito is tall. | (*C*) *P*(*Danny DeVito is tall*) is high. |

L&G argue that the paradox is resolved by treating premises and conclusions as probabilistic, rather than Boolean. This ties into the claim in their last sentence (top of the handout) that Bayesian models combine logical and probabilistic reasoning. Notice, however, that their model assumes Boolean truth-functional semantics, out-sourcing its probabilistic component into the interpretation process. Correspondingly, their 'solution' to the Sorites paradox depends on construing it in probabilistic (interpretive!) terms. Consequently, it 'solves' the Sorites paradox only to the extent that utterance *meaning* and *interpretation* are essentially the same thing.

***Note:*** On p. 604 L&G confusingly refer to the conclusion as 'the third premise'.