Lassiter and Goodman 2017: vagueness and RSA

- Lassiter, Daniel and Noah Goodman. 2017. Adjectival vagueness in a Bayesian model of interpretation. *Synthese* 194, 3801–3836. https://doi.org/10.1007/s11229-015-0786-1
  - Hallmarks of vagueness emerge from a probabilistic model of uptake.
- Barker, Chris. 2002. The dynamics of vagueness. *Linguistics and Philosophy* 25.1:1–36. doi:10.1023/a:1014346114955
  - "a use of a vague predicate affects the context against which other expressions get evaluated"
- Carter, Sam. To appear. Vagueness and discourse dynamics. *Linguistics meets Philosophy*. Cambridge University Press.
  - "Using [a vague expression] imposes contraints on future use."

Plan

1. Lassiter and Goodman 2017
2. I say how my priors have been updated by this seminar

Next session: student presentations

- Wed 5 May, 3:30 – 5:30
- Five presentations, each 20 minutes, including discussion
- Same link as the normal class; all welcome

Schedule:

1. Omar Agha
   - **The role of QUD-relevance in presupposition projection**
2. Nigel Flower
   - **The dynamics of presuppositional islands**
3. Zhuoye Zhao
   - **Supplement projection and discourse coherence**

5 minute break

4. Anna Alsop
   - **Hurford conjunctions in context**
5. Chris Barker
   - **Composing local contexts**

1. Lassiter and Goodman 2017

Some hallmarks of vagueness

- Context sensitivity
- Unclear criteria of application (“midtown”)
- Borderline cases
  - Some people are tall (positive extension)
  - Some people are not tall (negative extension)
  - Some people are neither tall nor not tall (borderline cases)
- Susceptibility to the Sorites paradox
  1. A person with a full head of hair is not bald.
  2. If someone isn’t bald, after removing one hair they still won’t be.
  3. Therefore a person with zero hairs is not bald.
- Tolerance
  - If a person is tall, then a person who is sufficiently close in height to that person is also tall
- Higher order vagueness

The dynamics of vagueness

- Constraints on future use
  - One can apply a vague predicate in a borderline case
  - Kyburg and Morreau 2000: two pigs, both borderline fat:
    - "I'll buy the fat one."
  - Lewis 1979:

Taking standards of precision as a component of conversational score, we once more find a rule of accommodation at work. One way to change the standards is to say something that would be unacceptable if the standards remained unchanged. If you say "Italy is boot-shaped" and get away with it, low standards are required and the standards fall if need be; thereafter "France is hexagonal" is true enough. But if you deny that Italy is boot-shaped, pointing out the differences, what you have said requires high standards under which "France is hexagonal" is far from true enough.

Figure 1: Lewis on standards of precision

- So standards for vague predicates go on the scoreboard
- Manage standards in the semantics? Or go pragmatic?

Lassiter and Goodman: vagueness for free

- No special-purpose machinery
- Following the lead of Edginton's view on vagueness
  - degree based
  - probabalistic
  - not credences, but "verities"
- assume *heavy* means *heavier than* θ
- "listeners must use the available information—context and knowledge of the speakers' beliefs and goals—to estimate this latent variable [i.e., θ]"
- profound question: how is communication possible in the presence of uncertainty arising from vagueness?

Preliminaries

- Enrichment of sets of worlds as models of propositions:
  - Probability measure P mapping sets of worlds into $[0,1]$
  - $P\emptyset = 0$
  - $PW = 1$
  - $P(A \cup B) = PA + PB$ for $A \cap B = \emptyset$
- Bayes' rule, ignoring the normalazing denominator:

$$P(B|A) \propto P(A|B) \times P(B)$$

RSA

- "listeners use their models of speakers' utterance choice to make more informed interpretive choices than would be possible if they simply updated their information states with the information that the utterance's semantic interpretation is true"
- If so, an update semantics is not fine-grained enough!
- $P_L(w|u) \propto P_S(u|w) \times P_L(w)$
  - "a listener L updates her information state, given that some utterance has been made, by reasoning about how the speaker would have chosen utterances or other actions in various possible worlds"
- $P_S(u|w) \propto P_L(w|u) \times P_S(u)$
  - "a speaker chooses utterances by reasoning about how the listener will interpret the utterance, together with some private utterance preferences $P_S(u)$" (e.g., brevity)

How many turtles?

- The pragmatic listener: L_1
  - L_1 reasons about the utterance choices of S_1
  - S_1 reasons about the intepretation choices of L_0
  - L_0 is the *literal listener*
- "the non-maximal speaker and listener models—S_1 and L_0—exist only as part of the pragmatic listener's psychology"
- Lewis ("Convention"): "we are windowless monads doing our best to mirror each other, mirror each other mirroring each other, and so on."

The literal listener, L_0, corresponds to dynamic semantics

- $P_{L_0}(w|u) = P_{L_0}(w|[\![u]\!] = 1)$
- "L_0 is essentially a probabilistic version of the interpreter discussed by Stalnaker (1978) and in much work in dynamic semantics"
- "There is even a close relationship between the update operations: Stalnakerian update is set intersection, and conditionalization is equivalent to intersection followed by renormalization of the measure"

S_1 attempts to address the QUD

- "In our model, the QUD provides a set of possible answers A over which the informativity of a potential utterance is calculated. A QUD might be "Who came to the party?", or "How many people came?", or "How tall is Al?""
- "The speaker and listener share the goal of coordinating utterance and interpretation so as to maximize the probability that the listener will infer the correct answer to the QUD."
- [We predict] "the utility of u for speaker S_1 to be proportional to its informativity to the literal listener L_0 about the true answer A, minus a non-negative cost $C(u)$."
- Remember, S_1 is a figment of L_1's rational mind. In order to model S_1, L_1 needs to estimate S_1's choices, which depends on S_1's goals.

Always pay attention to the cost

- "Relevant factors might include difficulty of articulation and difficulty of retrieval. Here we assume that cost increases monotonically with difficulty of articulation, which we approximate very roughly as length in words."

Utility = informativity − cost

- $\mathbb{U}_{S_1}(u; A) = ln(P_{L_0}(A|u)) - C(u)$
  - "$\mathbb{U}_{S_1}(u; A)$ is the utility of utterance u for speaker S_1 on the assumption that the answer to the QUD is A"
- That is, the value of an utterance to S_1 is probability that the literal listener will decide that A is the answer when hearing utterance u (minus the cost of uttering u).
- Two ways to increase utility:
  - reduce cost
  - make $P_{L_0}(A|u)$ bigger
    - reduce the number of other answers (B, C, ...) compatible with the truth of u

Soft-max

- Instead of assuming that agents choose the action with the highest utility, **soft-max**: speakers' likelihood of choosing an action is proportional to its utility
- $P_{S_1}(u|A) \propto exp(\alpha \times \mathbb{U}_{S_1}(u; A))$
  - "This choice rule has a parameter $\alpha > 0$ which determines how closely stochastic choice approximates deterministic utility-maximization. With $\alpha = \infty$, we would recover the choice rule typically used in game theory. In simulations reported below we set α to a lowish value of 4."
- α is λ in other discussions of RSA?

Alternatives

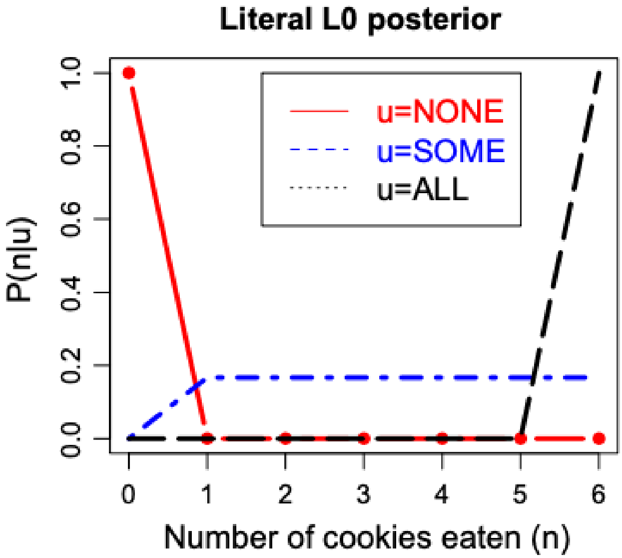
- Archimedes: “let me choose the set of alternatives and I can move the world”
- The choice of the set of relevant alternatives "can influence the qualitative behavior of the model, and it is not currently well-investigated from an empirical or computational perspective.
- For present purposes we will assume (essentially following Fox and Katzir (2011)) that the alternative utterances considered are a subset of the possible answers to the QUD.
- We will also assume that speakers have the option of saying nothing."
- “A major desideratum in future work will be to get a clearer picture of how speakers and listeners choose a realistic but manageable set of alternatives for pragmatic reasoning”

The pragmatic listener, L_1

- $P_{L_1}(A|u) \propto P_{S_1}(u|A) \times P_{L_1}(A)$
- "L_1 interprets utterances u using Bayesian inference, assigning to each [possible message A, intepreted as an answer to the QUD] a probability proportional to the product of
  - (a) the probability that the speaker would have chosen to employ u if A were the true answer, and
  - (b) the prior probability that A is true."
- "$P_{L_1}(A)$ specifies L_1's background knowledge about answers to the QUD.
  - For example, if the QUD is *How tall is Al?* and
  - L_1 knows only that Al is an adult man, then
  - $P_{L_1}(A)$ is an estimate of the distribution of heights among adult men"

Test drive: scalar implicatures

1. Charlie ate some of the cookies.



Estimating S_1

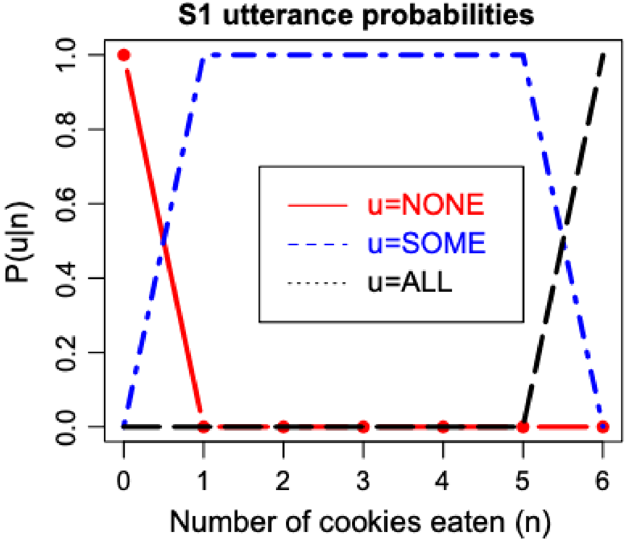


Figure 3: Posterior distrubtion for S_1 (.001 vs .999)

The pragmatic listener L_1 posterior

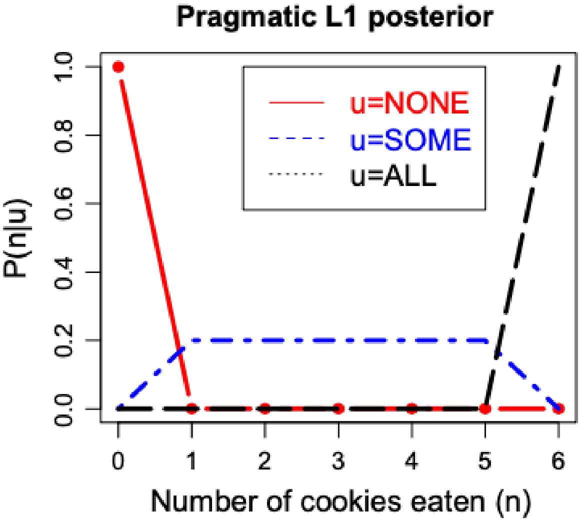


Figure 4: Posterior distrubtion for L_1

Analysis of the case in which $n = 6$

Now things get interesting: we have two true utterances in our alternative set, SOME and ALL.

- "There is a huge difference in informativity between SOME and ALL—the true state has probability six times higher if u = ALL than it does if u = SOME."

- $P_{S_1}(NONE|n = 6) \propto \exp(4 \times [\ln(0) - 4]) = 0.$
- $P_{S_1}(SOME|n = 6) \propto \exp(4 \times [\ln(1/6) - 4]) \approx 8.7 \times 10^{-11}.$
- $P_{S_1}(ALL|n = 6) \propto \exp(4 \times [\ln(1) - 4]) \approx 1.1 \times 10^{-7}.$

As a function of the informativity difference, S_1 is much more likely to produce *ALL* than *SOME* when *ALL* is true. The normalized probabilities are

- $P_{S_1}(NONE|n = 6) = 0.$
- $P_{S_1}(SOME|n = 6) \approx .001.$
- $P_{S_1}(ALL|n = 6) \approx .999.$

That is, a speaker who is trying to be informative *almost certainly* won't use *SOME* if they have the option to use *ALL* and *ALL* is true.[6] Using this information, the pragmatic

Figure 5: Now things get interesting

Upshot

- "The posterior probability of $n = 6$, given the observed utterance SOME, is much lower for the pragmatic listener L_1 than it is for a literal listener."
- "The difference is driven by the fact that the pragmatic listener considers not only what the speaker actually chose to say, but also other things that the speaker could have chosen."

Ok, adding sensitivity to thresholds

Let V be a function which assigns values to all variables in the language, and let $[\![u]\!]^V$ be the language's interpretation function as parametrized by V. The literal listener is as in the simple model, except that he conditions on $[\![u]\!]^V$, given a value of V which is provided by the speaker model.

(27) $$P_{L_0}(A|u, V) = P_{L_0}\big(A|[\![u]\!]^V = 1\big).$$

The variable-sensitive S_1 model also takes a given value of V and produces utterances stochastically, on the assumption that the variables are valued as they are in V.

(28) $$P_{S_1}(u|A, V) \propto \exp(\alpha \times \ln[P_{L_0}(A|u, V) - C(u)]).$$

The pragmatic listener then derives a variable-sensitive interpretation by considering how likely it is that the speaker would have said u if the answer were A *and* the variables were as in V—and, as usual, multiplying this value by the prior probabilities of A and V. This gives us a function which assigns a joint posterior probability to all possible combinations of A and V.[10]

(29) $$P_{L_1}(A, V|u) \propto P_{S_1}(u|A, V) \times P_{L_1}(A) \times P_{L_1}(V).$$

Figure 6: Assuming A and V are independent in L_1's prior

The literal meaning of gradable adjectives (Kennedy style)

(25)

a. $[\![tall]\!] = \lambda\theta_{tall}\lambda x[\mu_{tall}(x) > \theta_{tall}]$

b. $[\![POS]\!] = \lambda A\lambda x\lambda\theta_A[A(\theta_A)(x)]$

c. $[\![POS\ tall]\!] = \lambda x\lambda\theta_{tall}[\mu_{tall}(x) > \theta_{tall}]$

d. $[\![Al\ is\ POS\ tall]\!] = \lambda\theta_{tall}[\mu_{tall}(\mathbf{Al}) > \theta_{tall}]$.

Figure 7: Al is tall

Simulated posterior

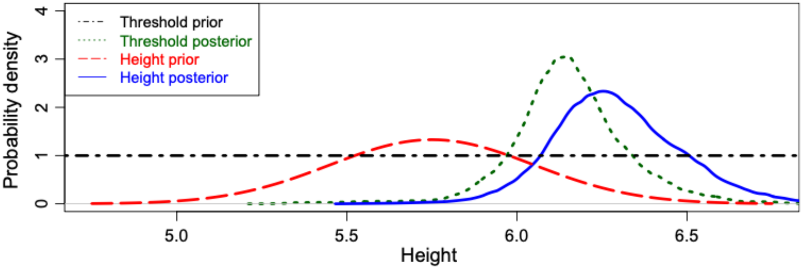


Figure 8: Priors and posterior

"The interpretation of a vague adjective in our model is a function from priors to posteriors"

- Interpretation, not literal denotation

Explanations of vague behavior

1. Context sensitivity: the interpretation process is highly sensitive to the form of the input prior.
2. Borderline cases: borderline cases of *tall* are individuals whose probability of counting as *tall* is intermediate.

(32) $$P_T(a \text{ is tall}) = \int_0^{height(a)} P_{L_1}(\theta_{tall} | u = \text{“}a \text{ is tall”})\, d\theta_{tall}.$$

The intuition behind Eq. 32 is this. If we know an individual's height and want to know how appropriate it would be to describe him as “tall”, we imagine ourselves using “tall” in communicating with a listener with an appropriate prior distribution $P_{L_{0/1}}(h)$. Using

Figure 9: How often do you see an integral in a semantics paper?

- This is “the probability that the utterance is true relative to the context-sensitive posterior on θ_{tall} that this L_1 derives”, NOT a theory of whether *Al is tall* is true.
- I.e., this is a theory of loose talk, of begin ‘true enough’—rather, of being informative enough

Explanations cont'd: sorites

(35)
a. x_n is tall.
b. If x_n is tall, then x_{n-1} is tall.
c. If x_{n-1} is tall, then x_{n-2} is tall.
d. If x_{n-2} is tall, then x_{n-3} is tall.
e. ...
f. If x_2 is tall, then x_1 is tall.
g. If x_1 is tall, then x_0 is tall.
h. $\therefore$ x_0 is tall.

Figure 10: Edginton's preferred form of the sorites inference

- Strategy: compute the probability of each premise from (a) through (g)

Premise probability

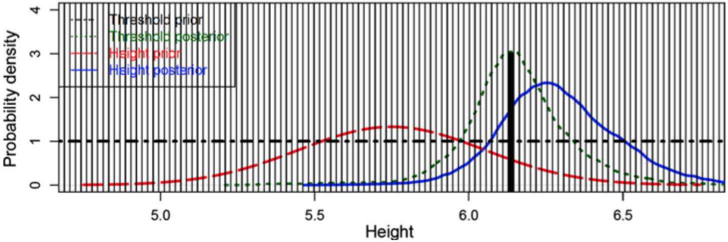


Figure 11: One line per premise; $\epsilon = .5$ inch

- Each narrow vertical rectangle corresponds to one premise
- The probability of the premise being true is equal to the probability that the threshold does not fall within that rectangle
- That probability is lowest in the filled in rectangle
- Yet the probability even for this rectangle is 97%
- By allowing the increment to go to zero, the minimal probability of the premises will approach certainty

Comments

- RSA is a quantitative model of update:
  - Given a QUD and a set of relevant alternative utterances,
  - Interpretation depends on literal meaning, informativity and cost
- As a side effect, evaluation-relevant priors get updated
- That is, the scoreboard gets adjusted
- Lassiter and Goodman: "The interpretation of a vague adjective in our model is a function from priors to posteriors"

2. How my priors have been updated by this seminar

Seminar recap

Starting point:

1. The value of expressions depends on context
   - extensions depend on the world of evaluation
   - anaphora/binding depends on operative assignment function
   - the *scoreboard*: a record of the evaluation-relevant aspects of the context

2. The use of an expression can be causally involved in updating the scoreboard
    - **Presupposition**: The content of some expressions can satisfy the presuppositions of other expressions
    - **Anaphora**: The use of an indefinite guarantees the availability of an antecedent for subsequent anaphora
        - within certain systematic limits (scope of negation, etc.)
    - **Modality**: The content of earlier expressions constrains the epistemic accessibility relation
    - **Vagueness**, **Loose talk**: use of expressions sets standards
3. At least some updating of the scoreboard is not grammatical
    - goats
    - accommodation

Big questions

- Is any context update systematic enough to be grammatical?
- If so, is a grammatical description the right explanation?
- Paradigm answers:
  - Heim 1983: expressions denote context update functions
  - Harris 2020: expressions don't even denote propositions
  - Mandelkern 2021:
    - "There is something dynamic about natural language...
    - ...but it is not truth conditions."
  - Schlenker 2007:
    - "something evolves dynamically in the course of a conversation.
    - But we claim that the only information that needs to be updated concerns *the words that the speech act participants have pronounced*."
  - Barker 2021 (next week):
    - The only information that needs to be updated concerns the semantic commitments of the expressions that have already been evaluated.

The internal/external distinction

- Is a dynamic semantics necessary in order to get truth conditions right?
  - A workspace in which the initial scoreboard gets temporarily updated
  - Purely in the service of computing truth conditions
  - **Internally dynamic**
  - Paradigm example: Charlow
    - To get cumulative readings right, track update of sets of assignments
- Do we want a dynamic semantics to tell us how to update the discourse context?
  - Belief manipulation (common ground)
  - List of salient antecedents
  - **Externally dynamic**
  - Paradigm examples: some expressions trigger “non-negotiable” updates to the discourse
    - Evidentials (Murray)
    - Appositives (ABH)
    - Imperatives (Starr)

- In favor of an externally dynamic semantics:
    - Heim, DPL, GSV, Murray, ABH, Starr, Carter: yes
    - Some expressions denote context change functions
    - Carter?: getting truth conditions right depends on tracking and modifying prevailing standards of loose talk
- Against an externally dynamic semantics:
    - Schlenker, Lewis, Mandelkern, me
    - Lewis:
        - discourse anaphora is plan recognition
        - plan recognition is not grammar
        - therefore discourse anaphora is not grammatical

My priors as of 2008:

At the end of the version of this seminar that took place in 2008, in the last class I presented a detailed recapituation, along with a statement of my views as reflecting what I learned from running the seminar.

- Is meaning dynamic? Hell, yeah! The order in which expressions are evaluated makes a detectable difference independent of the contribution to truth conditions.

- Do meanings update context? That would be nuts: meanings don't have the authority to change contexts. Words don't change contexts, people do.

Pryor's question: is there any way of saying what it means to be dynamic without making assumptions about the nature of the context (sets of worlds, assignment functions, etc.)?

Figure 12: Views from 2008

Some bets from 2008:

Truth conditions are static. Scorekeeping (especially anaphora) is dynamic.

Presuppositions and appositives trigger obligatory update; at-issue entailments are proferred content, and may or may not induce update.

Rendering the meaning of a sentence as a context update function does not guarantee that the function will be deployed. Sentences have a context update *potential*, not a context update effect.

Interpretation and update is incremental, but not governed by individual lexical items.

Figure 13: Views from 2008

My current view (FWIW)

- Something about natural language is dynamic
- Natural language is internally dynamic
- Natural language is not externally dynamic
- Expressions denote functions from points (world/assignment) to extensions (e.g., truth values)
- We must track (at least; perhaps only) semantic commitments
- The relevant sense of order is evaluation order
  - Often the same as linear order
  - Sometimes different (quantifier scope, reconstruction)
- Neutral about bivalence and classical logic

Bets

- Tracking semantic commitments is the right way to understand
  - presupposition satisfaction
  - bounded modality
- We will probably need to reconize multi-dimensional content
  - at-issue/NAI
  - evidential
  - expressive/appositive
  - neutral wrt questions of what's dynamic
- Anaphora... don't know

See you in 2034!

THANKS to you all!