

# Updating Alternatives in Pragmatic Competition

Sunwoo Jeong\* and James N. Collins†

Princeton University\* and the University of Hawai'i at Mānoa†

*sunwooj@princeton.edu and jamesnc@hawaii.edu*

March 15, 2019

# Overview

① What's an alternative?

② Experiment

③ Discussion

④ Conclusion

# Alternatives

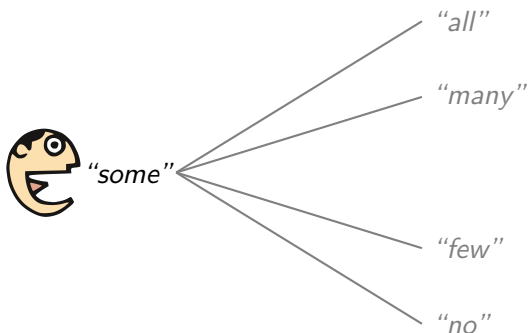
- How do interlocutors calculate a speaker's intended meaning given an underspecified literal meaning?

# Alternatives

- How do interlocutors calculate a speaker's intended meaning given an underspecified literal meaning?
- Since Grice 1975, a central component of this process is understood to be *alternatives*: expressions the speaker could have used.

# Alternatives

- How do interlocutors calculate a speaker's intended meaning given an underspecified literal meaning?
- Since Grice 1975, a central component of this process is understood to be *alternatives*: expressions the speaker could have used.



# The basic recipe

- A Gricean inference (an abbreviated “basic recipe” from Geurts 2009):

- (1)
  - a. Assume: The speaker utters “some”.
  - b. Assume: The speaker is cooperative.
  - c. The alternative “all” is more informative than “some”.
  - d. By (b) and (c), the speaker must lack evidence to assert “all”
  - e. Assuming the speaker is knowledgeable, she lacks evidence because “all” is false.

# The basic recipe

- A Gricean inference (an abbreviated “basic recipe” from Geurts 2009):
  - (1)
    - a. Assume: The speaker utters “some”.
    - b. Assume: The speaker is cooperative.
    - c. The alternative “all” is more informative than “some”.
    - d. By (b) and (c), the speaker must lack evidence to assert “all”
    - e. Assuming the speaker is knowledgeable, she lacks evidence because “all” is false.
- But why did we pick “all” in (c) as opposed to some other expression?

# The symmetry problem

- Kroch 1972: if we choose “some but not all” as the relevant alternative, the *opposite inference emerges*.



# The symmetry problem

- Kroch 1972: if we choose “some but not all” as the relevant alternative, the *opposite inference emerges*.
- (2)
- a. Assume: The speaker utters “some”.
  - b. Assume: The speaker is cooperative.
  - c. The alternative “some but not all” is more informative than “some”.
  - d. By (b) and (c), the speaker must lack evidence to assert “some but not all”
  - e. Assuming the speaker is knowledgeable, she lacks evidence because “some but not all” is false.
  - f. “some” conjoined with “not(some but not all)” is **all**

# The symmetry problem

- Kroch 1972: if we choose “some but not all” as the relevant alternative, the *opposite inference emerges*.
- (2)
  - a. Assume: The speaker utters “some”.
  - b. Assume: The speaker is cooperative.
  - c. The alternative “some but not all” is more informative than “some”.
  - d. By (b) and (c), the speaker must lack evidence to assert “some but not all”
  - e. Assuming the speaker is knowledgeable, she lacks evidence because “some but not all” is false.
  - f. “some” conjoined with “not(some but not all)” is **all**
- For a Gricean theory to be non-contradictory, we need some principled reason why **all** is an alternative but **some but not all** isn't.

# Lexicalized alternatives

- The neo-Gricean solution (Horn 1972, Gazdar 1979, Atlas and Levinson 1981 etc.): alternatives are lexicalized.

(3)

|       |   |
|-------|---|
| PHON: | "some"  |
| CAT:  | Det (DP/NP)   |
| SEM:  | $\{ \langle A, B \rangle \mid A \cap B \neq \emptyset \}$   |
| ALTS: | $\langle \textit{few}, \textit{many}, \textit{all} \rangle$ |

# Lexicalized alternatives

- The neo-Gricean solution (Horn 1972, Gazdar 1979, Atlas and Levinson 1981 etc.): alternatives are lexicalized.
- Horn and Abbott 2012: evidence for alternative scales comes from paradigmatic contrastive expressions.

(3)

|       |   |
|-------|---|
| PHON: | "some"  |
| CAT:  | Det (DP/NP)   |
| SEM:  | $\{ \langle A, B \rangle \mid A \cap B \neq \emptyset \}$   |
| ALTS: | $\langle \textit{few}, \textit{many}, \textit{all} \rangle$ |

- *not only X but Y*
- *X if not Y*
- *X or even Y*
- *X in fact Y*
- *not even X, much less Y*

# Lexicalized alternatives

- The neo-Gricean solution (Horn 1972, Gazdar 1979, Atlas and Levinson 1981 etc.): alternatives are lexicalized.
- Horn and Abbott 2012: evidence for alternative scales comes from paradigmatic contrastive expressions.

(3)

|       |   |
|-------|---|
| PHON: | "some"  |
| CAT:  | Det (DP/NP)   |
| SEM:  | $\{ \langle A, B \rangle \mid A \cap B \neq \emptyset \}$   |
| ALTS: | $\langle \textit{few}, \textit{many}, \textit{all} \rangle$ |

- *not only X but Y*
- *X if not Y*
- *X or even Y*
- *X in fact Y*
- *not even X, much less Y*

A theory which hard-codes alternatives via lexicalization need a way of verifying when and how items are lexicalized as alternatives.

# Structural approaches

- Katzir 2011: alternatives aren't lexicalized. An expression can compete with *any* expression of the same syntactic category.

## Structurally defined alternatives

The alternatives of a sentence  $S$  is any  $S'$  derived from  $S$  by:

- deleting nodes or,
- substituting lexical items

# Structural approaches

- Katzir 2011: alternatives aren't lexicalized. An expression can compete with *any* expression of the same syntactic category.

## Structurally defined alternatives

The alternatives of a sentence  $S$  is any  $S'$  derived from  $S$  by:

- deleting nodes or,
- substituting lexical items

- (4)
- a. ***Some*** of the students left.
  - b. ***All*** of the students left.
  - c. ***Some but not all*** of the students left.
- (b) is an alternative to (a) as it is derived by lexical substitution.
  - (c) is not an alternative as we have to insert extra material.

# Cost-based approaches

- An intuition from Grice: speakers prefer less complex expressions.
- e.g., Bergen et al 2016: *some but not all* is less preferred to *all* because of its structural complexity.

- (5)
- a. ***Some*** of the students left.
  - b. ***All*** of the students left.
  - c. ***Some but not all*** of the students left.

- The alternative (c) not ruled out; but the competition from (c) dampened because it is a more complex expression.



# Cost-based approaches

- An intuition from Grice: speakers prefer less complex expressions.
- e.g., Bergen et al 2016: *some but not all* is less preferred to *all* because of its structural complexity.

- (5)
- a. ***Some*** of the students left.
  - b. ***All*** of the students left.
  - c. ***Some but not all*** of the students left.

- The alternative (c) not ruled out; but the competition from (c) dampened because it is a more complex expression.

## Cost (Potts et al. 2016)

$C : M \mapsto \mathbb{R}$  is a cost function on messages.

For lexical items, **costs are specified**. For a non-terminal node  $A$  with daughters  $B_1 \dots B_n$ ,  $C(A) = \sum_{i=1}^n C(B_i)$ .

# At what cost?

- Our goal today: delve deeper into this notion of **cost**.

## Our guiding intuition about cost

An expression  $X$ 's cost reflects its “ease of use”, determined by several factors including structural complexity (e.g., frequency, politeness).

- Our study focuses on the relevance of an expression's *frequency* in the immediate discourse.
- More frequently used expressions should be “easier to use”, and thus have lower cost.

## Key hypothesis

$Y$  should implicate  $\neg \llbracket X \rrbracket$  more strongly each time  $X$  is used in the immediate discourse.

# Motivation & general design

Testing the hypothesis experimentally in the domain of epistemic modals.  
(see also: Schuster & Degen 2018, Lassiter 2016)

(6) It {**might** | **will** | **is likely to**} rain.

- *might* competes with more informative modals *will/likely*, implicating lower probabilities.
- This implicature should become stronger the more times the alternative expressions are used in the interaction.

# Motivation & general design

Testing the hypothesis experimentally in the domain of epistemic modals.  
(see also: Schuster & Degen 2018, Lassiter 2016)

(6) It {**might** | **will** | **is likely to**} rain.

- *might* competes with more informative modals *will/likely*, implicating lower probabilities.
- This implicature should become stronger the more times the alternative expressions are used in the interaction.

**The main task**     Rating the naturalness of a modal statement given contexts that vary in likelihood of rain

# Stimuli

Weather report with chance of rain in increments of 10%:

*0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%*

# Stimuli

Weather report with chance of rain in increments of 10%:

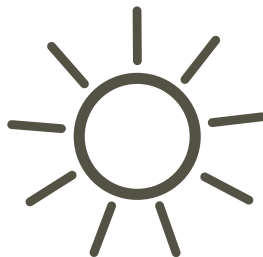
*0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%*

## WEATHER FORECAST

CHANCE OF RAIN:



20%



80%

# Stimuli

Weather report with chance of rain in increments of 10%:

*0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%*

WEATHER FORECAST

CHANCE OF RAIN:



80%

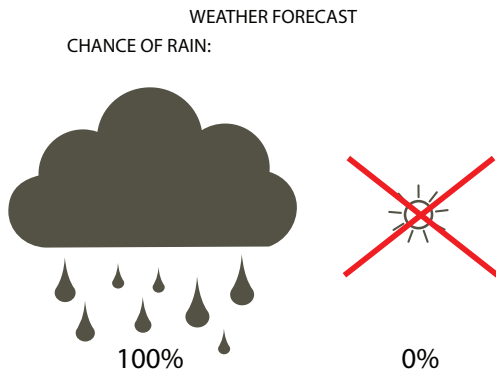


20%

# Stimuli

Weather report with chance of rain in increments of 10%:

*0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%*





# Stimuli

**Between-subject condition**      Different range of alternatives:

- (7)      **Condition without 'likely':** It {**might** | **will**} rain.
- a.    *might*: 3 times
  - b.    *will*: 3 times
- (8)      **Condition with 'likely':** It {**might** | **will** | **is likely to**} rain.
- a.    *might*: 2 times
  - b.    *will*: 2 times
  - c.    *be likely to*: 2 times

# Stimuli

**Between-subject condition**      Different range of alternatives:

- (7)      **Condition without 'likely':** It {**might** | **will**} rain.
- a.    *might*: 3 times
  - b.    *will*: 3 times
- (8)      **Condition with 'likely':** It {**might** | **will** | **is likely to**} rain.
- a.    *might*: 2 times
  - b.    *will*: 2 times
  - c.    *be likely to*: 2 times

**For condition with 'likely'**      Also tracked # of 'likely' encountered up to the current trial:

# Procedure

Lily is a firm believer of the weather forecast and thinks that it's always accurate. Mr. Tansley asks Lily what the weather will be like tomorrow. Lily remembers that the forecast for tomorrow was as follows:

## WEATHER FORECAST

CHANCE OF RAIN:



40%



60%

Lily tells Mr. Tansley:

"It might rain."

- Questions in a trial:
  - **Q1:** Given what Lily knows, is her statement above true or false? (forced choice)
  - **Q2:** How naturally does Lily's utterance describe the state of the world? (ratings from 0–100 on a slider bar)
- 10 trials: 6 target trials, 4 fillers/controls
- Target trials paired with 6 different contexts (pseudo-randomized)

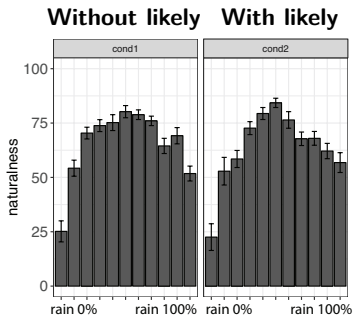
# Procedure

- **Participants** 480 native speakers of American English from Amazon Mechanical Turk
- **Analysis** A series of mixed effects regression models fitted to *might* data, with:
  - Naturalness as the main dependent variable
  - (i) context, (ii) condition or *likely* count as predictors
  - interaction between the two above
  - Random intercepts for participants

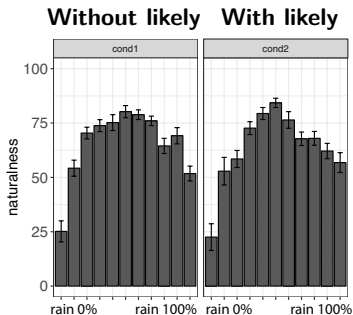
# Procedure

- **Participants** 480 native speakers of American English from Amazon Mechanical Turk
- **Analysis** A series of mixed effects regression models fitted to *might* data, with:
  - Naturalness as the main dependent variable
  - (i) context, (ii) condition or *likely* count as predictors
  - interaction between the two above
  - Random intercepts for participants
- **Data** Here we focus solely on *might* trials

# Results: condition effect

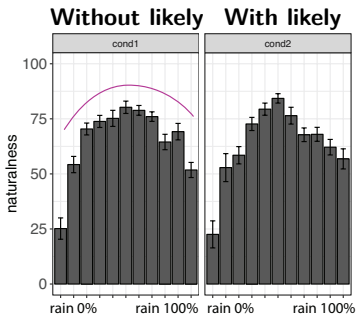


# Results: condition effect



- Naturalness of *might* across 2 conditions

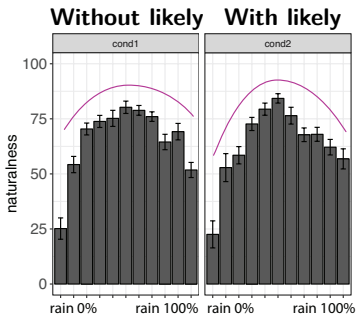
# Results: condition effect



- Naturalness of *might* across 2 conditions

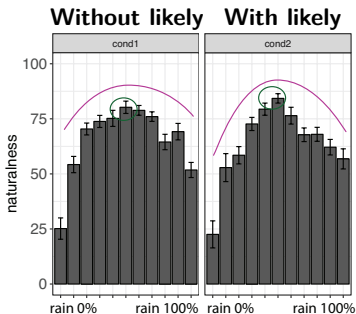


# Results: condition effect



- Naturalness of *might* across 2 conditions

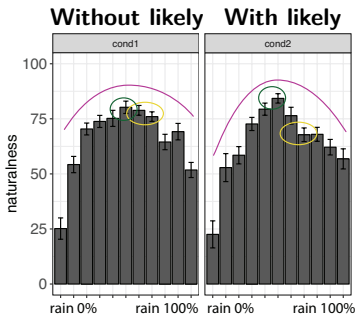
# Results: condition effect



$$\beta = -27.10, S.E. = 3.71, t = -7.3, **$$

- Naturalness of *might* across 2 conditions
- *might* significantly less natural in 80–100% region in the 'likely' condition

# Results: condition effect

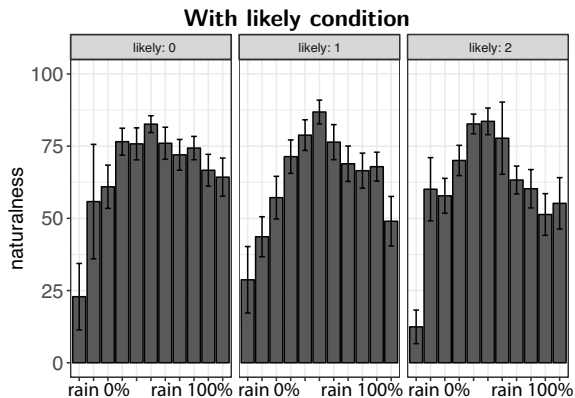


$$\beta = -27.10, S.E. = 3.71, t = -7.3, **$$

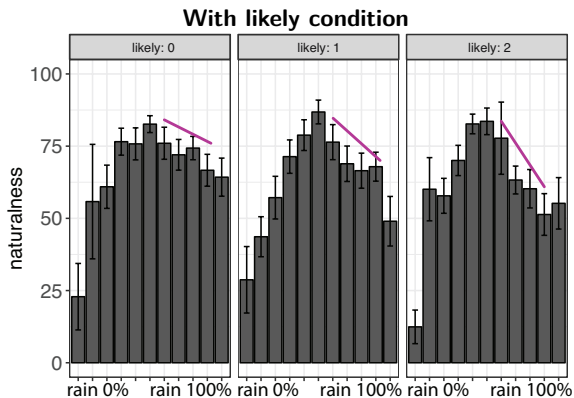
$$\beta = -25.73, S.E. = 4.22, t = -6.09, **$$

- Naturalness of *might* across 2 conditions
- *might* significantly less natural in 80–100% region in the ‘likely’ condition
- *might* significantly less natural in 0–20% region in the ‘likely’ condition

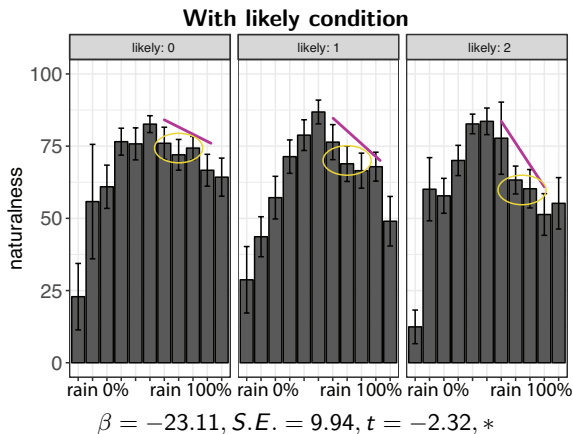
## Results: frequency effect



# Results: frequency effect

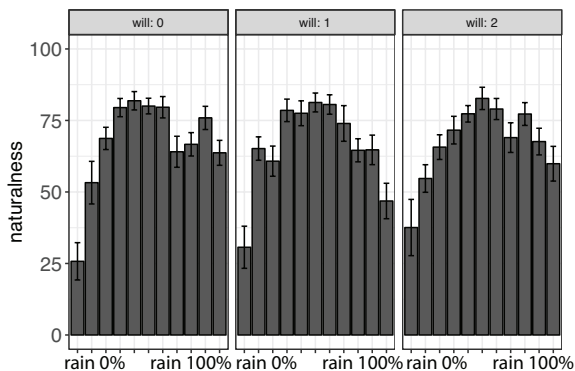


# Results: frequency effect

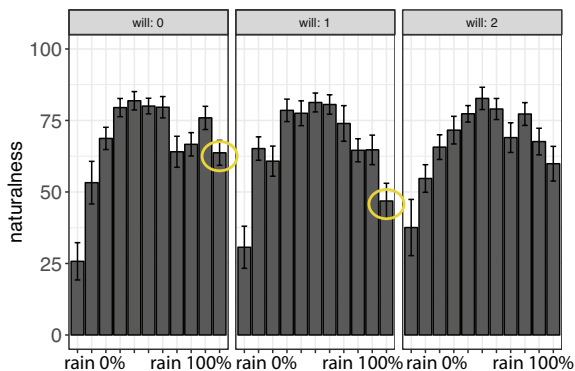


- *might* worse in 70–100% region the more one encounters *likely*

# Results: frequency effect – will?



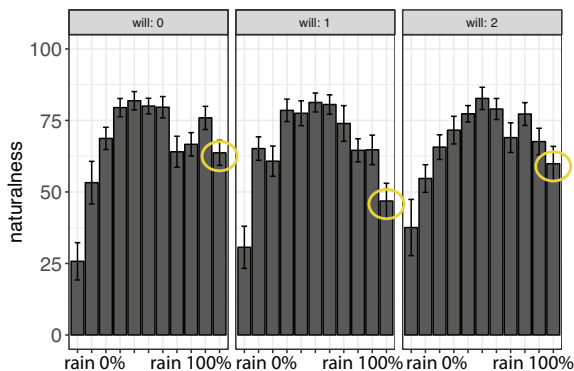
# Results: frequency effect – will?



- *might* worse in 100% region after encountering *will* once



# Results: frequency effect – will?



- But naturalness of *might* in 100% region goes up again after encountering *will* twice

# Summary of results

- Predictions broadly confirmed & hypothesis corroborated
- The implicature  $\neg\llbracket\textit{likely}\rrbracket$  is strengthened the more one encounters *likely*

## Key hypothesis

$Y$  should implicate  $\neg\llbracket X\rrbracket$  more strongly each time  $X$  is used in the immediate discourse.

# Summary of results

- Predictions broadly confirmed & hypothesis corroborated
- The implicature  $\neg \llbracket \textit{likely} \rrbracket$  is strengthened the more one encounters *likely*

## Key hypothesis

$Y$  should implicate  $\neg \llbracket X \rrbracket$  more strongly each time  $X$  is used in the immediate discourse.

- The result suggests a model whereby listeners incorporate information about frequency into their pragmatic reasoning.

# Cost in pragmatic theory

- Our notion of the *cost* of  $X$ : “ease of use” of  $X$ .
- How do we incorporate this into pragmatic theory?

# Cost in pragmatic theory

- Our notion of the *cost* of  $X$ : “ease of use” of  $X$ .
- How do we incorporate this into pragmatic theory?

## The ‘speaker’ in RSA (Lassiter and Goodman 2017)

The speaker weights preferences between alternatives based on *utility* ( $\mathbb{U}$ ).

$$\mathbb{U}il_{S_1}(uttr.|Answ, \theta) = \ln(LitListnr(Ans|uttr, \theta) - Cost(uttr))$$

- The speaker weighs
  - i. the likelihood the listener will choose answer  $A$  given utterance  $u$  and contextual standard  $\theta$ .
  - ii. **the cost of uttering  $u$ .**

# What is cost?

- We propose to articulate several parameters entering into the calculation of cost of  $u$ :
  - ① The structural complexity of  $u$  (cf. Katzir 2011, Potts et al. 2016).
  - ② The politeness/social appropriateness of  $u$  (cf. Yoon et al 2016).
  - ③ The baseline frequency of  $u$ .
  - ④ How recent was the last occurrence of  $u$
  - ⑤ The frequency of  $u$  in the immediate discourse.
  - ⑥ ...

# What is cost?

- We propose to articulate several parameters entering into the calculation of cost of  $u$ :
  - ① The structural complexity of  $u$  (cf. Katzir 2011, Potts et al. 2016).
  - ② The politeness/social appropriateness of  $u$  (cf. Yoon et al 2016).
  - ③ The baseline frequency of  $u$ .
  - ④ How recent was the last occurrence of  $u$
  - ⑤ The frequency of  $u$  in the immediate discourse.
  - ⑥ ...

## Unpacking cost

$$C(u) = \text{Freq}(u) \cdot \text{Complex}(u) \cdot \text{Polite}(u) \cdot \text{Rec}(u) \cdot \dots$$

- A priori, *might* is unlikely to compete with *indubitably* due to its baseline low frequency ( $\therefore$  high  $C$ ).
- But if a speaker demonstrates a willingness to use *indubitably* ( $\therefore$  lower  $C$ ), it should compete with *might*.

# Discourse frequency and cost

- Our primary focus: *the frequency of  $u$  in the immediate discourse*.
- $DFreq(u)$  'the discourse frequency of  $u$ ': a parameter which lowers cost each time  $u$  is encountered in the discourse.

## Discourse Frequency

$$DFreq(u) = \exp(-\frac{n}{\tau})$$

$n$  — the no. times  $u$  has been used in the immediate discourse,  
 $\tau$  — a sensitivity parameter.



# Discourse frequency and cost

- Our primary focus: *the frequency of  $u$  in the immediate discourse*.
- $DFreq(u)$  'the discourse frequency of  $u$ ': a parameter which lowers cost each time  $u$  is encountered in the discourse.

## Discourse Frequency

$$DFreq(u) = \exp(-\frac{n}{\tau})$$

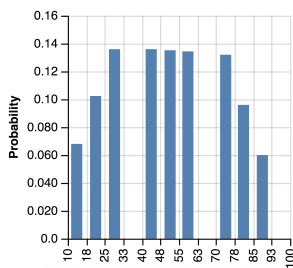
$n$  — the no. times  $u$  has been used in the immediate discourse,  
 $\tau$  — a sensitivity parameter.

- Let  $\tau = 6$ .  $DFreq('likely')$  lowers as  $n$  increases.
- The baseline cost of 'likely' may be lowered when multiplied by  $DFreq('likely')$  depending on the value of  $n$ .

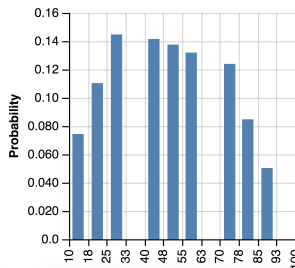
- (9)
- Cond1:**  $DFreq('likely') = \exp(-0/6) = 1$
  - Cond2:**  $DFreq('likely') = \exp(-1/6) = 0.846$
  - Cond3:**  $DFreq('likely') = \exp(-2/6) = 0.717$

# A simulation in RSA

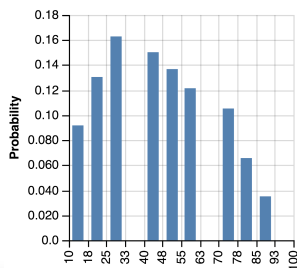
- The effect is demonstrated using RSA (Lassiter and Goodman 2017).
  - $\llbracket \text{might}(\text{rain}) \rrbracket = 1$  iff  $P(\text{rain!}) > 0$
  - $\llbracket \text{likely}(\text{rain}) \rrbracket = 1$  iff  $P(\text{rain!}) > \theta$
- The likelihood  $L_1$  assigns to each chance of rain given an utterance of *might*. *likely* becomes a better competitor each time it is used.
  - Assuming flat priors on  $\theta$  and normal distribution over rain likelihood.



$n(\text{'likely'}) = 0$



$n(\text{'likely'}) = 1$



$n(\text{'likely'}) = 2$

# Conclusion

- The big question: what are the constraints and factors that determine relevant alternatives in pragmatic inferences?
- Established one factor: interlocutors' willing to use an alternative in a given discourse, signalled by frequency in the interaction.
- Pragmatic competition sensitive to a host of contextual factors, including metalinguistic factors like the *ease of use* a form.

# Selected References



Lassiter, D. and N. D. Goodman (2017)

Adjectival vagueness in a Bayesian model of interpretation

Synthese 194: 3801–3836.



Potts, C., et al. (2016)

Embedded implicatures as pragmatic inferences under compositional lexical uncertainty

Journal of Semantics 33: 755–802.



Schuster, S. and J. Degen. (2018).

Adaptation to variable use of expressions of uncertainty

Poster presented at AMLaP 2018.