SEMINAR

A Practical Introduction to The Rational Speech Act Modeling Framework

Lecturer: Tan Zhi Xuan
Student: Martin Nguyen (Duc Q. Nguyen)

Department of Computer Science
National University of Singapore

September 17^{th} 2025

"Most people do not listen with the intent to understand; they listen with the intent to reply."

Stephen R. Covey

Table of Contents

Table of Contents

Motivation

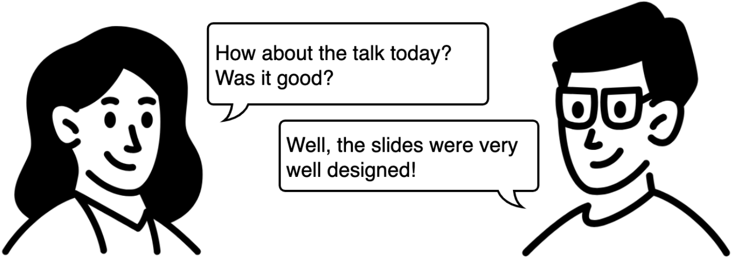
Warm-up question

Do you think the girl has accepted the invitation?

Motivation

Question 1

Suppose you run into your friend who just attended a talk, so ask them if it was good. Your friend says, "Well, the slides were very well designed!"
What would you infer about the talk? Would a literal listener make the same inference? And how would you model the goals / utility of your friend as the speaker, in order to explain the inference you made?



Language as Social Bayesian Inference

Motivation:

- **Rigid literal semantics** misses context and speaker intent.
- In practice, interpretation depends on **almost everything** in context.
- Listeners reason about *how* speakers choose utterances, and speakers reason about *how* listeners interpret their utterances.

[1] Gregory Scontras, Tessler, and Franke, *A practical introduction to the Rational Speech Act modeling framework*.

Language as Social Bayesian Inference

Motivation:

- **Rigid literal semantics** misses context and speaker intent.
- In practice, interpretation depends on **almost everything** in context.
- Listeners reason about *how* speakers choose utterances, and speakers reason about *how* listeners interpret their utterances.

Core idea: Make pragmatics explicit and probabilistic:

Meaning $\approx$ Bayesian inference over beliefs, goals, and costs.

[1] Gregory Scontras, Tessler, and Franke, *A practical introduction to the Rational Speech Act modeling framework*.

Language as Social Bayesian Inference

Motivation:

- **Rigid literal semantics** misses context and speaker intent.
- In practice, interpretation depends on **almost everything** in context.
- Listeners reason about *how* speakers choose utterances, and speakers reason about *how* listeners interpret their utterances.

Core idea: Make pragmatics explicit and probabilistic:

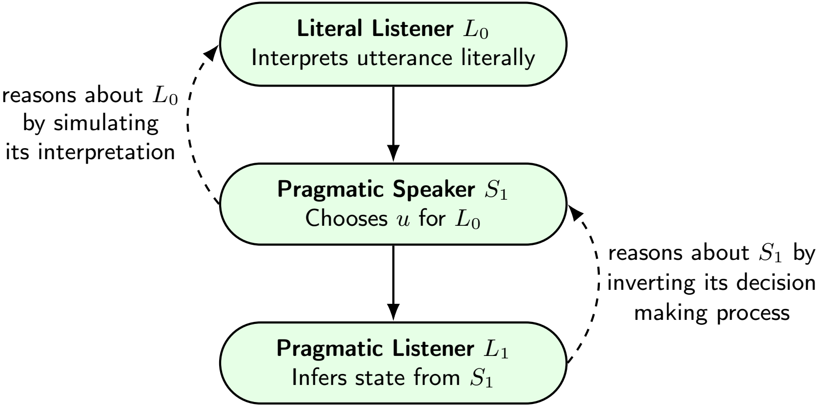
Meaning $\approx$ Bayesian inference over beliefs, goals, and costs.

What is the Rational Speech Act Modeling Framework?

Rational Speech Act (RSA) modeling is a framework for capturing how speakers and listeners interact by recursively reasoning about each other's interpretations of utterances, starting from their literal meanings.

[1] Gregory Scontras, Tessler, and Franke, *A practical introduction to the Rational Speech Act modeling framework*.

Vanilla RSA



Literal Listener L_0
Interprets utterance literally
Pragmatic Speaker S_1
Chooses u for L_0
Pragmatic Listener L_1
Infers state from S_1
reasons about L_0 by simulating its interpretation
reasons about S_1 by inverting its decision making process

Vanilla RSA

Literal listener (L_0): Interprets the utterance using only its literal semantics and prior beliefs.

$$P_{L_0}(s \mid u) \propto P(s) \llbracket u \rrbracket (s)$$

[1] **Gregory Scontras, Tessler, and Franke, *A practical introduction to the Rational Speech Act modeling framework*.**

Vanilla RSA

Literal listener (L_0): Interprets the utterance using only its literal semantics and prior beliefs.

$$P_{L_0}(s \mid u) \propto P(s)\, [\![u]\!](s)$$

Pragmatic speaker (S_1): Chooses an utterance that maximizes informativeness for L_0 while minimizing cost.

$$P_{S_1}(u \mid s) \propto \exp\Big(\alpha\big[\log P_{L_0}(s \mid u) - C(u)\big]\Big)$$

[1] **Gregory Scontras, Tessler, and Franke, *A practical introduction to the Rational Speech Act modeling framework*.**

Vanilla RSA

Literal listener (L_0): Interprets the utterance using only its literal semantics and prior beliefs.

$$P_{L_0}(s \mid u) \propto P(s)\, [\![u]\!](s)$$

Pragmatic speaker (S_1): Chooses an utterance that maximizes informativeness for L_0 while minimizing cost.

$$P_{S_1}(u \mid s) \propto \exp\Big(\alpha \big[\log P_{L_0}(s \mid u) - C(u)\big]\Big)$$

Pragmatic listener (L_1): Infers the intended state by reasoning about why the speaker chose that utterance.

$$P_{L_1}(s \mid u) \propto P(s)\, P_{S_1}(u \mid s),$$

where α: speaker optimality; $C(u)$: utterance cost; $[\![u]\!]$: literal truth-conditions.

[1] **Gregory Scontras, Tessler, and Franke, *A practical introduction to the Rational Speech Act modeling framework*.**

Vanilla RSA

Literal listener (L_0): Interprets the utterance using only its literal semantics and prior beliefs.

$$P_{L_0}(s \mid u) \propto P(s)\, [\![u]\!](s)$$

Pragmatic speaker (S_1): Chooses an utterance that maximizes informativeness for L_0 while minimizing cost.

$$P_{S_1}(u \mid s) \propto \exp\Big(\alpha\big[\log P_{L_0}(s \mid u) - C(u)\big]\Big)$$

Pragmatic listener (L_1): Infers the intended state by reasoning about why the speaker chose that utterance.

$$P_{L_1}(s \mid u) \propto P(s)\, P_{S_1}(u \mid s),$$

where α: speaker optimality; $C(u)$: utterance cost; $[\![u]\!]$: literal truth-conditions.

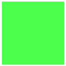
Side question

What do you think if we start with a literal speaker?

[1] **Gregory Scontras, Tessler, and Franke, *A practical introduction to the Rational Speech Act modeling framework*.**

Example: Referential Communication Game

World states (uniform prior $P(s_i) = \frac{1}{3}$):

$s_1 =$ , $s_2 =$ , $s_3 =$ .

Utterances: $u \in \{\text{blue}, \text{green}, \text{square}, \text{circle}\}$.

Literal semantics $[\![u]\!](s) \in \{0, 1\}$:

$$[\![\text{blue}]\!] : \{s_1, s_2\}, \quad [\![\text{green}]\!] : \{s_3\},$$
$$[\![\text{square}]\!] : \{s_1, s_3\}, \quad [\![\text{circle}]\!] : \{s_2\}.$$

[1] Gregory Scontras, Tessler, and Franke, *A practical introduction to the Rational Speech Act modeling framework*.

Literal Listener L_0

$$P_{L_0}(s \mid u) \;\propto\; P(s)\,[\![u]\!](s) \quad \text{then normalize over } s.$$

Example: Referential Communication Game

Literal Listener L_0

$$P_{L_0}(s \mid u) \;\propto\; P(s)\,[\![u]\!](s) \quad \text{then normalize over } s.$$

(a) Hearing $u =$ **blue** ($[\![\text{blue}]\!] = \{s_1, s_2\}$):

$$P_{L_0}(s_1 \mid \text{blue}) = \tfrac{1}{2},\; P_{L_0}(s_2 \mid \text{blue}) = \tfrac{1}{2},\; P_{L_0}(s_3 \mid \text{blue}) = 0.$$

Example: Referential Communication Game

Literal Listener L_0

$$P_{L_0}(s \mid u) \;\propto\; P(s)\,[\![u]\!](s) \quad \text{then normalize over } s.$$

(a) Hearing $u =$ **blue** ($[\![\text{blue}]\!] = \{s_1, s_2\}$):

$$P_{L_0}(s_1 \mid \text{blue}) = \tfrac{1}{2},\; P_{L_0}(s_2 \mid \text{blue}) = \tfrac{1}{2},\; P_{L_0}(s_3 \mid \text{blue}) = 0.$$

(b) Hearing $u =$ **green** ($[\![\text{green}]\!] = \{s_3\}$):

$$P_{L_0}(s_1 \mid \text{green}) = 0, \quad P_{L_0}(s_2 \mid \text{green}) = 0, \quad P_{L_0}(s_3 \mid \text{green}) = 1.$$

Example: Referential Communication Game

Literal Listener L_0

$$P_{L_0}(s \mid u) \ \propto \ P(s)\,[\![u]\!](s) \quad \text{then normalize over } s.$$

(a) Hearing $u =$ **blue** ($[\![\text{blue}]\!] = \{s_1, s_2\}$):

$$P_{L_0}(s_1 \mid \text{blue}) = \tfrac{1}{2},\ P_{L_0}(s_2 \mid \text{blue}) = \tfrac{1}{2},\ P_{L_0}(s_3 \mid \text{blue}) = 0.$$

(b) Hearing $u =$ **green** ($[\![\text{green}]\!] = \{s_3\}$):

$$P_{L_0}(s_1 \mid \text{green}) = 0, \quad P_{L_0}(s_2 \mid \text{green}) = 0, \quad P_{L_0}(s_3 \mid \text{green}) = 1.$$

(c) Hearing $u =$ **square** ($[\![\text{square}]\!] = \{s_1, s_3\}$):

$$P_{L_0}(s_1 \mid \text{square}) = \tfrac{1}{2},\ P_{L_0}(s_3 \mid \text{square}) = \tfrac{1}{2},\ P_{L_0}(s_2 \mid \text{square}) = 0.$$

Example: Referential Communication Game

Literal Listener L_0

$$P_{L_0}(s \mid u) \;\propto\; P(s)\,[\![u]\!](s) \quad \text{then normalize over } s.$$

(a) Hearing $u =$ **blue** ($[\![\text{blue}]\!] = \{s_1, s_2\}$):

$$P_{L_0}(s_1 \mid \text{blue}) = \tfrac{1}{2},\; P_{L_0}(s_2 \mid \text{blue}) = \tfrac{1}{2},\; P_{L_0}(s_3 \mid \text{blue}) = 0.$$

(b) Hearing $u =$ **green** ($[\![\text{green}]\!] = \{s_3\}$):

$$P_{L_0}(s_1 \mid \text{green}) = 0, \quad P_{L_0}(s_2 \mid \text{green}) = 0, \quad P_{L_0}(s_3 \mid \text{green}) = 1.$$

(c) Hearing $u =$ **square** ($[\![\text{square}]\!] = \{s_1, s_3\}$):

$$P_{L_0}(s_1 \mid \text{square}) = \tfrac{1}{2},\; P_{L_0}(s_3 \mid \text{square}) = \tfrac{1}{2},\; P_{L_0}(s_2 \mid \text{square}) = 0.$$

(d) Hearing $u =$ **circle** ($[\![\text{circle}]\!] = \{s_2\}$):

$$P_{L_0}(s_1 \mid \text{circle}) = 0, \quad P_{L_0}(s_3 \mid \text{circle}) = 0, \quad P_{L_0}(s_2 \mid \text{circle}) = 1.$$

Example: Referential Communication Game

Pragmatic Speaker S_1 with $\alpha = 1,\ C(u) = 0$

$$P_{S_1}(u \mid s) \ \propto\ \exp\big(\log P_{L_0}(s \mid u)\big) \ =\ P_{L_0}(s \mid u).$$

Thus S_1 is proportional to the corresponding L_0 column and then *row-normalized* over u.

Example: Referential Communication Game

Pragmatic Speaker S_1 with $\alpha = 1,\ C(u) = 0$

$$P_{S_1}(u \mid s) \ \propto\ \exp\big(\log P_{L_0}(s \mid u)\big) \ = \ P_{L_0}(s \mid u).$$

Thus S_1 is proportional to the corresponding L_0 column and then *row-normalized* over u.

(a) For $s_1 =$ **blue square:** non-zero L_0: blue : $\frac{1}{2}$, square : $\frac{1}{2}$.

$$\{0.5, 0.5\} \text{ normalized } \Rightarrow P_{S_1}(\text{blue} \mid s_1) = 0.5, \quad P_{S_1}(\text{square} \mid s_1) = 0.5.$$

Example: Referential Communication Game

Pragmatic Speaker S_1 with $\alpha = 1,\ C(u) = 0$

$$P_{S_1}(u \mid s)\ \propto\ \exp\big(\log P_{L_0}(s \mid u)\big)\ =\ P_{L_0}(s \mid u).$$

Thus S_1 is proportional to the corresponding L_0 column and then *row-normalized* over u.

(a) For $s_1 =$ blue square: non-zero L_0: blue : $\frac{1}{2}$, square : $\frac{1}{2}$.

$$\{0.5, 0.5\} \text{ normalized } \Rightarrow P_{S_1}(\text{blue} \mid s_1) = 0.5, \quad P_{S_1}(\text{square} \mid s_1) = 0.5.$$

(b) For $s_2 =$ blue circle: non-zero L_0: blue : $\frac{1}{2}$, circle : 1.

$$\{0.5, 1\} \text{ normalized } \Rightarrow P_{S_1}(\text{blue} \mid s_2) = \tfrac{1}{3} \approx 0.33, \quad P_{S_1}(\text{circle} \mid s_2) = \tfrac{2}{3} \approx 0.67.$$

Example: Referential Communication Game

Pragmatic Speaker S_1 with $\alpha = 1,\ C(u) = 0$

$$P_{S_1}(u \mid s) \ \propto\ \exp\big(\log P_{L_0}(s \mid u)\big) \ =\ P_{L_0}(s \mid u).$$

Thus S_1 is proportional to the corresponding L_0 column and then *row-normalized* over u.

(a) For s_1 = blue square: non-zero L_0: blue : $\frac{1}{2}$, square : $\frac{1}{2}$.

$$\{0.5, 0.5\} \text{ normalized } \Rightarrow P_{S_1}(\text{blue} \mid s_1) = 0.5, \quad P_{S_1}(\text{square} \mid s_1) = 0.5.$$

(b) For s_2 = blue circle: non-zero L_0: blue : $\frac{1}{2}$, circle : 1.

$$\{0.5, 1\} \text{ normalized } \Rightarrow P_{S_1}(\text{blue} \mid s_2) = \tfrac{1}{3} \approx 0.33, \quad P_{S_1}(\text{circle} \mid s_2) = \tfrac{2}{3} \approx 0.67.$$

(c) For s_3 = green square: non-zero L_0: green : 1, square : $\frac{1}{2}$.

$$\{1, 0.5\} \text{ normalized } \Rightarrow P_{S_1}(\text{green} \mid s_3) = \tfrac{2}{3} \approx 0.67, \quad P_{S_1}(\text{square} \mid s_3) = \tfrac{1}{3} \approx 0.33.$$

Example: Referential Communication Game

Pragmatic Listener L_1

$$P_{L_1}(s \mid u) \;\propto\; P(s)\, P_{S_1}(u \mid s) \quad \text{then normalize over } s.$$

Example: Referential Communication Game

Pragmatic Listener L_1

$$P_{L_1}(s \mid u) \;\propto\; P(s)\, P_{S_1}(u \mid s) \quad \text{then normalize over } s.$$

(a) Hearing $u =$ blue:

$$\tilde{P}(s_1) = \tfrac{1}{3} \cdot 0.5 = \tfrac{1}{6} \approx 0.1667, \quad \tilde{P}(s_2) = \tfrac{1}{3} \cdot 0.33 \approx 0.11, \quad \tilde{P}(s_3) = 0.$$

Normalize by $0.1667 + 0.11 \approx 0.2767$:

$$P_{L_1}(s_1 \mid \text{blue}) \approx 0.60, \quad P_{L_1}(s_2 \mid \text{blue}) \approx 0.40, \quad P_{L_1}(s_3 \mid \text{blue}) = 0.$$

Example: Referential Communication Game

Pragmatic Listener L_1

$$P_{L_1}(s \mid u) \;\propto\; P(s)\, P_{S_1}(u \mid s) \quad \text{then normalize over } s.$$

(a) Hearing $u =$ blue:

$$\tilde{P}(s_1) = \tfrac{1}{3} \cdot 0.5 = \tfrac{1}{6} \approx 0.1667, \quad \tilde{P}(s_2) = \tfrac{1}{3} \cdot 0.33 \approx 0.11, \quad \tilde{P}(s_3) = 0.$$

Normalize by $0.1667 + 0.11 \approx 0.2767$:

$$P_{L_1}(s_1 \mid \text{blue}) \approx 0.60, \quad P_{L_1}(s_2 \mid \text{blue}) \approx 0.40, \quad P_{L_1}(s_3 \mid \text{blue}) = 0.$$

(b) Hearing $u =$ green: only s_3 has mass $\Rightarrow P_{L_1}(s_3 \mid \text{green}) = 1$, others 0.

Pragmatic Listener L_1

$$P_{L_1}(s \mid u) \;\propto\; P(s)\, P_{S_1}(u \mid s) \quad \text{then normalize over } s.$$

(a) Hearing $u =$ **blue:**

$$\tilde{P}(s_1) = \tfrac{1}{3} \cdot 0.5 = \tfrac{1}{6} \approx 0.1667, \quad \tilde{P}(s_2) = \tfrac{1}{3} \cdot 0.33 \approx 0.11, \quad \tilde{P}(s_3) = 0.$$

Normalize by $0.1667 + 0.11 \approx 0.2767$:

$$P_{L_1}(s_1 \mid \text{blue}) \approx 0.60, \quad P_{L_1}(s_2 \mid \text{blue}) \approx 0.40, \quad P_{L_1}(s_3 \mid \text{blue}) = 0.$$

(b) Hearing $u =$ **green:** only s_3 has mass $\Rightarrow P_{L_1}(s_3 \mid \text{green}) = 1$, others 0.
(c) Hearing $u =$ **square:**

$$\tilde{P}(s_1) = \tfrac{1}{3} \cdot 0.5 = \tfrac{1}{6} \approx 0.1667, \quad \tilde{P}(s_2) = 0, \quad \tilde{P}(s_3) = \tfrac{1}{3} \cdot 0.33 \approx 0.11,$$

Normalize by 0.2767:

$$P_{L_1}(s_1 \mid \text{square}) \approx 0.60, \quad P_{L_1}(s_2 \mid \text{square}) = 0, \quad P_{L_1}(s_3 \mid \text{square}) \approx 0.40.$$

Example: Referential Communication Game

Pragmatic Listener L_1

$$P_{L_1}(s \mid u) \;\propto\; P(s)\, P_{S_1}(u \mid s) \quad \text{then normalize over } s.$$

(a) Hearing $u =$ **blue:**

$$\tilde{P}(s_1) = \tfrac{1}{3} \cdot 0.5 = \tfrac{1}{6} \approx 0.1667, \quad \tilde{P}(s_2) = \tfrac{1}{3} \cdot 0.33 \approx 0.11, \quad \tilde{P}(s_3) = 0.$$

Normalize by $0.1667 + 0.11 \approx 0.2767$:

$$P_{L_1}(s_1 \mid \text{blue}) \approx 0.60, \quad P_{L_1}(s_2 \mid \text{blue}) \approx 0.40, \quad P_{L_1}(s_3 \mid \text{blue}) = 0.$$

(b) Hearing $u =$ **green:** only s_3 has mass $\Rightarrow P_{L_1}(s_3 \mid \text{green}) = 1$, others 0.
(c) Hearing $u =$ **square:**

$$\tilde{P}(s_1) = \tfrac{1}{3} \cdot 0.5 = \tfrac{1}{6} \approx 0.1667, \quad \tilde{P}(s_2) = 0, \quad \tilde{P}(s_3) = \tfrac{1}{3} \cdot 0.33 \approx 0.11,$$

Normalize by 0.2767:

$$P_{L_1}(s_1 \mid \text{square}) \approx 0.60, \quad P_{L_1}(s_2 \mid \text{square}) = 0, \quad P_{L_1}(s_3 \mid \text{square}) \approx 0.40.$$

(d) Hearing $u =$ **circle:** only s_2 has mass $\Rightarrow P_{L_1}(s_2 \mid \text{circle}) = 1$, others 0.

Example: Referential Communication Game

World states with uniform prior:

Literal listener L_0

			
blue	0.5	0.5	0
green	0	0	1
square	0.5	0	0.5
circle	0	1	0

Pragmatic speaker S_1

	blue	green	square	circle
	0.5	0	0.5	0
	0.33	0	0	0.67
	0	0.67	0.33	0

Pragmatic listener L_1

			
blue	0.6	0.4	0
green	0	0	1
square	0.6	0	0.4
circle	0	1	0

[1] Gregory Scontras, Tessler, and Franke, *A practical introduction to the Rational Speech Act modeling framework*.

Informativity from Alignment of Beliefs

Motivation: Vanilla RSA defines speaker utility as:

$$U_{S_1}(u; s) = \log P_{L_0}(s \mid u) - C(u).$$

Side question

Why is the utility function in this form?

Informativity from Alignment of Beliefs

Motivation: Vanilla RSA defines speaker utility as:

$$U_{S_1}(u; s) = \log P_{L_0}(s \mid u) - C(u).$$

Side question

Why is the utility function in this form?

It can be derived from the idea that speakers aim to **align listener beliefs with their own**.

Informativity from Alignment of Beliefs

Motivation: Vanilla RSA defines speaker utility as:

$$U_{S_1}(u; s) = \log P_{L_0}(s \mid u) - C(u).$$

Side question

Why is the utility function in this form?

It can be derived from the idea that speakers aim to **align listener beliefs with their own**.

KL-Divergence Perspective

- Speaker has belief distribution $P_{S_1}(s)$ over states.
- Listener after hearing u: $P_{L_0}(s|u)$.
- Alignment measured by **Kullback–Leibler divergence**:

$$\mathsf{KL}(P_{S_1}(s) \parallel P_{L_0}(s|u)) = \sum_s P_{S_1}(s) \log \frac{P_{S_1}(s)}{P_{L_0}(s|u)}.$$

- Utility: minimize KL + cost:

$$U(u; s) = -\mathsf{KL}(P_{S_1}(s) \parallel P_{L_0}(s|u)) - C(u).$$

Informativity from Alignment of Beliefs

Motivation: Vanilla RSA defines speaker utility as:

$$U_{S_1}(u;s) = \log P_{L_0}(s \mid u) - C(u).$$

Why is the utility function in this form? – It can be derived from the idea that speakers aim to **align listener beliefs with their own**.

Special Case: Vanilla RSA

If speaker knows true state s^* ($P_{S_1}(s)$ is degenerate):

$$P_{S_1}(s^*) = 1 \qquad P_{S_1}(s') = 0, \forall s' \neq s^*$$

$$\mathsf{KL}(P_{S_1}(s) \,\|\, P_{L_0}(s \mid u)) = \log \frac{1}{P_{L_0}(s^* \mid u)} + \sum_{s \neq s^*} 0. \log \frac{0}{P_{L_0}(s \mid u)} = -\log P_{L_0}(s^* \mid u)$$

$$\Rightarrow \quad U(u;s) = \log P_{L_0}(s^* \mid u) - C(u).$$

[1] **Gregory Scontras, Tessler, and Franke,** ***A practical introduction to the Rational Speech Act modeling framework.***

Informativity from Alignment of Beliefs

Motivation: Vanilla RSA defines speaker utility as:

$$U_{S_1}(u; s) = \log P_{L_0}(s \mid u) - C(u).$$

Why is the utility function in this form? – It can be derived from the idea that speakers aim to **align listener beliefs with their own**.

Special Case: Vanilla RSA

If speaker knows true state s^* ($P_{S_1}(s)$ is degenerate):

$$P_{S_1}(s^*) = 1 \qquad P_{S_1}(s') = 0, \forall s' \neq s^*$$

$$\mathsf{KL}(P_{S_1}(s) \,\|\, P_{L_0}(s \mid u)) = \log \frac{1}{P_{L_0}(s^* \mid u)} + \sum_{s \neq s^*} 0. \log \frac{0}{P_{L_0}(s \mid u)} = -\log P_{L_0}(s^* \mid u)$$

$$\Rightarrow \quad U(u; s) = \log P_{L_0}(s^* \mid u) - C(u).$$

Side question

In practice, how can we deal with $\log(0)$?

[1] Gregory Scontras, Tessler, and Franke, *A practical introduction to the Rational Speech Act modeling framework*.

Relation to The Assistant Game

Question 2

Similar to the iterative best-response strategy for assistance games (where a principal models an assistant modeling a principal, etc.), RSA involves a listener recursively modeling a speaker (or vice versa, a speaker recursively modeling a listener). Who do you think best corresponds to the principal – the speaker, or the listener? And how do the two differ in their goals?

Vanilla RSA assumes:

- **Fixed semantics**: Shared, context-independent meaning for utterances.
- **Single goal**: Speakers aim to communicate the true state efficiently.
- **Fixed context**: Both agents share the same background assumptions.

[1] Gregory Scontras, Tessler, and Franke, *A practical introduction to the Rational Speech Act modeling framework*.

Extending the RSA Model

Vanilla RSA assumes:

- **Fixed semantics**: Shared, context-independent meaning for utterances.
- **Single goal**: Speakers aim to communicate the true state efficiently.
- **Fixed context**: Both agents share the same background assumptions.

However, real-world communication involves uncertainties:

- **Semantic ambiguous**: e.g., vague adjectives like *heavy*.
- **Figurative language**: e.g., hyperbole, metaphor.
- **Contextual factors**: e.g., comparison class for adjectives.
- **Speaker's epistemic state**: e.g., partial knowledge.
- **Speaker's utility calculus**: e.g., politeness vs. informativeness.

Each extension modifies the speaker's utility function or the listener's inference to account for these uncertainties.

[1] Gregory Scontras, Tessler, and Franke, *A practical introduction to the Rational Speech Act modeling framework*.

Table of Contents

Meaning Inference

Problem: Uncertainty about word meanings.

- Vanilla RSA assumes fixed semantics.
- In reality: words like *heavy*, *tall* depend on context.
- Speakers and listeners may not share the same lexicon.

[1] **Gregory Scontras, Tessler, and Franke, *A practical introduction to the Rational Speech Act modeling framework*.**

Meaning Inference

Problem: Uncertainty about word meanings.

- Vanilla RSA assumes fixed semantics.
- In reality: words like *heavy*, *tall* depend on context.
- Speakers and listeners may not share the same lexicon.

Key idea: Parameterize the interpretation function with a semantic variable x.

$$[[\text{heavy}]]_x = \{s : \text{weight}(s) > x\}$$

Revised RSA equations:

- **Pragmatic Speaker:**

$$P_{S_1}(u \mid s, x) \propto \exp\left(\alpha(\log P_{L_0}(s \mid u, x) - C(u))\right)$$

- **Pragmatic Listener:** Jointly infers state s and semantic parameter x:

$$P_{L_1}(s, x \mid u) \propto P_{S_1}(u \mid s, x) P(s) P(x)$$

Applications: Vague adjectives, generics, ambiguity resolution, scope ambiguity, lexical uncertainty.

[1] Gregory Scontras, Tessler, and Franke, ***A practical introduction to the Rational Speech Act modeling framework***.

Question Under Discussion Inference

Problem: Vanilla RSA assumes speakers always aim to communicate the true world state. In reality, speakers often have a specific *Question Under Discussion (QUD)*.

Example

"I paid a million dollars for coffee" $\rightarrow$ signals negative affect, not literal price.

[1] **Gregory Scontras, Tessler, and Franke,** ***A practical introduction to the Rational Speech Act modeling framework.***

Question Under Discussion Inference

Problem: Vanilla RSA assumes speakers always aim to communicate the true world state. In reality, speakers often have a specific *Question Under Discussion (QUD)*.

Example

"I paid a million dollars for coffee" $\rightarrow$ signals negative affect, not literal price.

Key idea: Introduce a QUD variable x that partitions the state space into cells (answers to the question).

$$F(s,x) = \text{answer to QUD } x \text{ given state } s$$

Revised RSA equations:

- **Pragmatic Speaker:**

$$P_{S_1}(u \mid s,x) \propto \exp\left(\alpha(\log P_{L_0}(F(s,x) \mid u) - C(u))\right)$$

- **Pragmatic Listener:** Infers both the world state s and which QUD x the speaker addresses:

$$P_{L_1}(s,x \mid u) \propto P_{S_1}(u \mid s,x)P(s)P(x)$$

Applications: Hyperbole, Irony, Metaphor

[1] Gregory Scontras, Tessler, and Franke, *A practical introduction to the Rational Speech Act modeling framework*.

Context / Prior Inference

Problem: Vanilla RSA assumes context is fixed and shared. In reality, listeners may not know the relevant comparison class or background assumptions.

Example

"He is tall" $\rightarrow$ tall compared to what? Basketball players or people in general?

[1] Gregory Scontras, Tessler, and Franke, *A practical introduction to the Rational Speech Act modeling framework*.

Context / Prior Inference

Problem: Vanilla RSA assumes context is fixed and shared. In reality, listeners may not know the relevant comparison class or background assumptions.

Example

"He is tall" $\rightarrow$ tall compared to what? Basketball players or people in general?

Key idea: Introduce a contextual variable x representing comparison class or background assumptions. The listener reasons jointly about:

- World state s
- Context x that the speaker had in mind

Revised RSA equations:

- **Pragmatic Speaker:** $P_{S_1}(u \mid s, x) \propto \exp\left(\alpha(\log P_{L_0}(s \mid u, x) - C(u))\right)$
- **Pragmatic Listener:** Infers both the state s and context x:

$$P_{L_1}(s, x \mid u) \propto P_{S_1}(u \mid s, x)\, P(s \mid x)\, P(x)$$

Applications: Comparison class inference, scalar implicatures under strong priors, and contextual adaptation.

[1] Gregory Scontras, Tessler, and Franke, *A practical introduction to the Rational Speech Act modeling framework*.

Epistemic Inference

Problem: Vanilla RSA assumes the speaker knows the true world state. In reality, speakers often have partial knowledge.

Example

"John ate some of the pizza." $\rightarrow$ If the speaker saw only crumbs, they don't know whether John ate all the pizza.

[1] Gregory Scontras, Tessler, and Franke, ***A practical introduction to the Rational Speech Act modeling framework***.

Epistemic Inference

Problem: Vanilla RSA assumes the speaker knows the true world state. In reality, speakers often have partial knowledge.

Example

"John ate some of the pizza." $\rightarrow$ If the speaker saw only crumbs, they don't know whether John ate all the pizza.

Key idea: Model the speaker's epistemic state x (their knowledge) and allow the listener to reason about it.

Revised RSA equations:

- **Pragmatic Speaker:**

$$P_{S_1}(u \mid x) \propto \exp\left(\alpha\, \mathbb{E}_{P(s|x)}[\log P_{L_0}(s \mid u)] - C(u)\right)$$

  - Expectation integrates over possible states consistent with x
- **Pragmatic Listener:** Infers both the world state s and the speaker's knowledge x

Applications: Scalar implicatures, plural predication (distributive vs. collective), competence inference (expert vs. novice).

[1] **Gregory Scontras, Tessler, and Franke, *A practical introduction to the Rational Speech Act modeling framework*.**

Complex Utility / Utility Inference

Problem: Vanilla RSA assumes speakers care only about informativeness. In reality, speakers often have social goals: politeness, saving face, or being kind.

Example

Instead of "Your haircut is ugly," someone says "It's an interesting look."

Key idea: Extend the speaker's utility to include multiple components:

- **Informational utility:** How well the utterance conveys the true state.
- **Social utility:** How positively the listener perceives the state (or the speaker).

Revised Speaker Rule:

$$P_{S_1}(u \mid s, x) \propto \exp\left(\alpha\left(x \cdot \log P_{L_0}(s \mid u) + (1-x) \cdot \mathbb{E}_{P_{L_0}(s|u)}[V(s)] - C(u)\right)\right)$$

- x: weight on informativeness ($0 \leq x \leq 1$); $(1-x)$: weight on social utility
- $V(s)$: subjective value of state s for the listener

Interpretation:

- $x \approx 1 \rightarrow$ speaker prioritizes informativeness
- $x \approx 0 \rightarrow$ speaker prioritizes social goals (politeness, compliments)

Applications: Politeness, white lies, goal inference (polite, sarcastic, informative).

Utterance Salience

Problem: Vanilla RSA models speaker cost $C(u)$ (effort) in the utility.

- Speakers are biased by **salience**: some utterances come to mind more easily.
- Modeling salience via an **utterance prior**, complementing or replacing cost.

Example

In US, people tend to use "floor" while in Singapore, "storey" is more common.

[1] **Gregory Scontras, Tessler, and Franke,** ***A practical introduction to the Rational Speech Act modeling framework.***

Utterance Salience

Problem: Vanilla RSA models speaker cost $C(u)$ (effort) in the utility.

- Speakers are biased by **salience**: some utterances come to mind more easily.
- Modeling salience via an **utterance prior**, complementing or replacing cost.

Example

In US, people tend to use "floor" while in Singapore, "storey" is more common.

Key idea: Add salience as a prior over utterances.

$$P_{S_1}^{\text{sal}}(u \mid s) \propto \text{Truth}(u; s) \cdot \text{Informativity}(u)^{\alpha} \cdot \underbrace{\text{Salience}(u)}_{P_{S_1}(u)}$$

- $\text{Salience}(u)$ = frequency, ease-of-retrieval, conventionality.
- Avoid double-counting: use salience prior, cost, or a principled combination.

Applications:

- **Psychological realism**: captures retrieval/frequency effects.
- **Explains**: overuse of short/common referring expressions.
- **NLP generation**: improves choice among near-equivalent utterances.

[1] Gregory Scontras, Tessler, and Franke, *A practical introduction to the Rational Speech Act modeling framework*.

Table of Contents

Key Ingredients & Challenges

- Explicit specification of:
  - **World states:** Which situations are possible?
  - **Utterances:** What expressions can be produced?
  - **Priors:** Speaker and listener beliefs over states/utterances.
- Numerical parameters to estimate: speaker optimality (α), utterance costs, decision thresholds.

[1] **Gregory Scontras, Tessler, and Franke,** ***A practical introduction to the Rational Speech Act modeling framework*****.**

Building Models with RSA

Key Ingredients & Challenges

- Explicit specification of:
  - **World states:** Which situations are possible?
  - **Utterances:** What expressions can be produced?
  - **Priors:** Speaker and listener beliefs over states/utterances.
- Numerical parameters to estimate: speaker optimality (α), utterance costs, decision thresholds.

Modeling Workflow

1. **Define spaces:** Specify states and utterances (via experimental design or pre-tests).
2. **Set priors:** Use empirical estimates or uniform baselines.
3. **Link to data:** Choose observation models (categorical, ordinal, continuous ratings).
4. **Estimate parameters:** Bayesian inference for α, costs, thresholds.
5. **Evaluate models:** Compare with others using Bayes Factors or predictive fit (LOO, WAIC).

[1] Gregory Scontras, Tessler, and Franke, *A practical introduction to the Rational Speech Act modeling framework*.

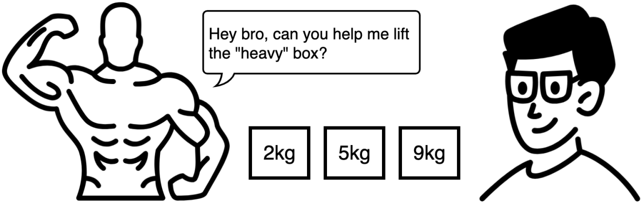
A More Complex Example

Imagine your friend, who is a gymer, is asking you to help lift a box. There are three boxes:

light (2kg), medium (5kg), heavy (9kg).

Your friend can describe the box weight with one of three utterances:

`say_light`, `say_medium`, `say_heavy`.



Challenge

Consider the case of your friend asking you to help lift a "heavy" box. Which box will you pick? Do you think the box is actually heavy?

A More Complex Example

Domain. States and weights:

$$\mathcal{S} = \{\text{light}(2),\ \text{medium}(5),\ \text{heavy}(9)\}.$$

Utterances: `say_light`, `say_medium`, `say_heavy`.

Key model features:

- States have a uniform prior distribution.
- Meaning inference only for `say_heavy`: threshold $x \in \{3, 7\}$.
- Speaker (your friend) incurs different costs for different utterances.
- Listener (you) has prior $P(x = 3) = 0.2,\ P(x = 7) = 0.8$ because you know your friend is a gymer and you can assume they have a higher threshold for "heavy".

Literal semantics and L_0

Literal truth-conditions:

$$\texttt{say\_light}: \ w \le 2,$$
$$\texttt{say\_medium}: \ 2 < w \le 5,$$
$$\texttt{say\_heavy}: \ w > x \quad (x \in \{3, 7\}).$$

Literal listener L_0:

$$P_{L_0}(s \mid u, x) \propto \mathbf{1}\{u \text{ true in } s\} \cdot P(s).$$

$\mathbf{x = 3}$			
Utterance	P(light)	P(medium)	P(heavy)
say_light	1.0	0.0	0.0
say_medium	0.0	1.0	0.0
say_heavy	0.0	0.5	0.5

$\mathbf{x = 7}$			
Utterance	P(light)	P(medium)	P(heavy)
say_light	1.0	0.0	0.0
say_medium	0.0	1.0	0.0
say_heavy	0.0	0.0	1.0

Speaker Model

Speaker utility:

$$U(u; s, x) = \log P_{L_0}(s \mid u, x) - C(u).$$

Speaker's utterance distribution:

$$P_{S_1}(u \mid s, x) \propto \exp(\alpha \cdot U(u; s, x)) \text{ , with } \alpha = 1.$$

Cost for utterance: $C(\text{say\_light}) = 1, \quad C(\text{say\_medium}) = 3, \quad C(\text{say\_heavy}) = 2.$

Speaker Model

Speaker utility:

$$U(u; s, x) = \log P_{L_0}(s \mid u, x) - C(u).$$

Speaker's utterance distribution:

$$P_{S_1}(u \mid s, x) \propto \exp(\alpha \cdot U(u; s, x))\,, \text{ with } \alpha = 1.$$

Cost for utterance: $C(\text{say\_light}) = 1, \quad C(\text{say\_medium}) = 3, \quad C(\text{say\_heavy}) = 2.$
$P_{S_1}(u \mid s, x = 3)$:

State	say_light	say_medium	say_heavy
light	1.00	0.00	0.00
medium	0.00	0.42	0.58
heavy	0.00	0.00	1.00

Speaker Model

Speaker utility:

$$U(u; s, x) = \log P_{L_0}(s \mid u, x) - C(u).$$

Speaker's utterance distribution:

$$P_{S_1}(u \mid s, x) \propto \exp(\alpha \cdot U(u; s, x)), \text{ with } \alpha = 1.$$

Cost for utterance: $C(\text{say\_light}) = 1, \quad C(\text{say\_medium}) = 3, \quad C(\text{say\_heavy}) = 2.$
$P_{S_1}(u \mid s, x = 3)$:

State	say_light	say_medium	say_heavy
light	1.00	0.00	0.00
medium	0.00	0.42	0.58
heavy	0.00	0.00	1.00

$P_{S_1}(u \mid s, x = 7)$:

State	say_light	say_medium	say_heavy
light	1.00	0.00	0.00
medium	0.00	1.00	0.00
heavy	0.00	0.00	1.00

Listener Inference

Pragmatic listener jointly infers state s and threshold x:

$$P_{L_1}(s, x \mid u) \propto P_{S_1}(u \mid s, x)\, P(s)\, P(x).$$

Listener Inference

Pragmatic listener jointly infers state s and threshold x:

$$P_{L_1}(s, x \mid u) \propto P_{S_1}(u \mid s, x)\, P(s)\, P(x).$$

$P_{L_1}(s, x \mid u = \text{say\_heavy})$

State	x=3	x=7
light	0.000	0.000
medium	0.103	0.000
heavy	0.179	0.717

$P_{L_1}(s, x \mid u = \text{say\_medium})$

State	x=3	x=7
light	0.000	0.000
medium	0.096	0.904
heavy	0.000	0.000

$P_{L_1}(s, x \mid u = \text{say\_light})$

State	x=3	x=7
light	0.2	0.8
medium	0.0	0.0
heavy	0.0	0.0

Listener Inference

Pragmatic listener jointly infers state s and threshold x:

$$P_{L_1}(s, x \mid u) \propto P_{S_1}(u \mid s, x)\, P(s)\, P(x).$$

$P_{L_1}(s, x \mid u = \text{say\_heavy})$

State	x=3	x=7
light	0.000	0.000
medium	0.103	0.000
heavy	0.179	0.717

$P_{L_1}(s, x \mid u = \text{say\_medium})$

State	x=3	x=7
light	0.000	0.000
medium	0.096	0.904
heavy	0.000	0.000

$P_{L_1}(s, x \mid u = \text{say\_light})$

State	x=3	x=7
light	0.2	0.8
medium	0.0	0.0
heavy	0.0	0.0

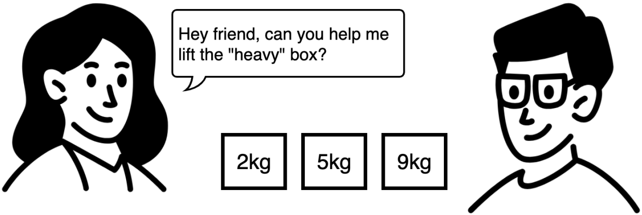
Insights

In case your gymer friend says "heavy" box, it is likely that the box is actually heavy. But, you might also think that the box is not that heavy because your friend likes to say the word "heavy" more than other words.

How about this context?

Side question

In case a girl asks you to lift a “heavy” box. Do you think it is actually heavy?

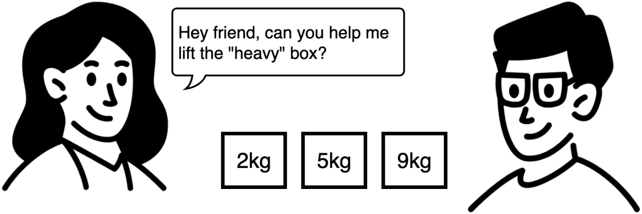


With the same setting, but now your prior is $P(x = 3) = P(x = 7) = 0.5$.

How about this context?

Side question

In case a girl asks you to lift a "heavy" box. Do you think it is actually heavy?



With the same setting, but now your prior is $P(x=3) = P(x=7) = 0.5$.

$$P_{L_1}(s, x \mid u = \text{say\_heavy})$$

State	x=3	x=7
light	0.000	0.000
medium	0.224	0.000
heavy	0.388	0.388

Playground

Playground Code: Colab Notebook

Relation to Large Language Models

Question 3

The paper covers many examples where people pragmatically understand others' language (e.g. understanding hyperbole, inferring whether "tall" means "tall for a child" vs. "tall for a basketballer", etc.). Do you think that LLMs have learned to make similar pragmatic inferences? If so, how do you think this was learned without explicit recursive agent modeling? If not, where do you think LLMs might fail, and how might you test that?

Table of Contents

Extensions Summary

Core Limitations of Vanilla RSA:

- Utterance-level, ignores **compositionality**.
- Static priors/lexica, no **adaptation**.
- Normative only, not a **processing model**.

[1] Gregory Scontras, Tessler, and Franke, *A practical introduction to the Rational Speech Act modeling framework*.

Extensions Summary

Core Limitations of Vanilla RSA:

- Utterance-level, ignores **compositionality**.
- Static priors/lexica, no **adaptation**.
- Normative only, not a **processing model**.

Key Extensions:

- **Compositionality:** integrate semantics via probabilistic programming.
- **Adaptation:** update priors $\Rightarrow$ learning & personalization.
- **Development:** captures child/adult word learning, ambiguity resolution, cultural evolution.
- **Processing Links:** rationality α, recursion depth, sampling, incremental RSA.

[1] **Gregory Scontras, Tessler, and Franke,** ***A practical introduction to the Rational Speech Act modeling framework*.**

Extensions Summary

Core Limitations of Vanilla RSA:

- Utterance-level, ignores **compositionality**.
- Static priors/lexica, no **adaptation**.
- Normative only, not a **processing model**.

Key Extensions:

- **Compositionality:** integrate semantics via probabilistic programming.
- **Adaptation:** update priors $\Rightarrow$ learning & personalization.
- **Development:** captures child/adult word learning, ambiguity resolution, cultural evolution.
- **Processing Links:** rationality α, recursion depth, sampling, incremental RSA.

Applications in NLP:

- **Learning semantics** via pragmatic reasoning.
- **Pragmatic wrappers** for neural models (captioning, navigation, MT, summarization).

[1] Gregory Scontras, Tessler, and Franke, *A practical introduction to the Rational Speech Act modeling framework*.

Extensions Summary

Core Limitations of Vanilla RSA:

- Utterance-level, ignores **compositionality**.
- Static priors/lexica, no **adaptation**.
- Normative only, not a **processing model**.

Key Extensions:

- **Compositionality:** integrate semantics via probabilistic programming.
- **Adaptation:** update priors $\Rightarrow$ learning & personalization.
- **Development:** captures child/adult word learning, ambiguity resolution, cultural evolution.
- **Processing Links:** rationality α, recursion depth, sampling, incremental RSA.

Applications in NLP:

- **Learning semantics** via pragmatic reasoning.
- **Pragmatic wrappers** for neural models (captioning, navigation, MT, summarization).

Takeaway: RSA serves as a **unified framework** for semantics & pragmatics, learning, processing, and NLP.

[1] Gregory Scontras, Tessler, and Franke, *A practical introduction to the Rational Speech Act modeling framework*.

Compositionality: Why It Matters

Limitation of Vanilla RSA:

- Operates at the **utterance level**: assumes a fixed, pre-composed meaning.
- Ignores how meanings are **built compositionally** from lexical items and syntax.
- Cannot fully capture phenomena like:
  - Quantifier scope
  - Embedded implicatures
  - Vagueness in complex structures

[1] N. D. Goodman and Lassiter, "Probabilistic Semantics and Pragmatics Uncertainty in Language and Thought".
[2] N. D. Goodman and Andreas, *The Design and Implementation of Probabilistic Programming Languages*.

Compositionality: Why It Matters

Limitation of Vanilla RSA:

- Operates at the **utterance level**: assumes a fixed, pre-composed meaning.
- Ignores how meanings are **built compositionally** from lexical items and syntax.
- Cannot fully capture phenomena like:
  - Quantifier scope
  - Embedded implicatures
  - Vagueness in complex structures

Proposed Extensions:

- Integrate **compositional semantics** into RSA reasoning chain.
- Use **probabilistic programming** for:
  - Stochastic lambda calculus[1]
  - Semantic parsing inside L_0[2]

[1] N. D. Goodman and Lassiter, "Probabilistic Semantics and Pragmatics Uncertainty in Language and Thought".
[2] N. D. Goodman and Andreas, *The Design and Implementation of Probabilistic Programming Languages*.

Compositionality: Why It Matters

Limitation of Vanilla RSA:

- Operates at the **utterance level**: assumes a fixed, pre-composed meaning.
- Ignores how meanings are **built compositionally** from lexical items and syntax.
- Cannot fully capture phenomena like:
  - Quantifier scope
  - Embedded implicatures
  - Vagueness in complex structures

Proposed Extensions:

- Integrate **compositional semantics** into RSA reasoning chain.
- Use **probabilistic programming** for:
  - Stochastic lambda calculus[1]
  - Semantic parsing inside L_0[2]

Benefits:

- Reduces ad-hoc specification of alternatives.
- Enables dynamic generation of utterances and verifying worlds.
- Brings RSA closer to a **unified theory of semantics & pragmatics**.

[1] N. D. Goodman and Lassiter, "Probabilistic Semantics and Pragmatics Uncertainty in Language and Thought".
[2] N. D. Goodman and Andreas, *The Design and Implementation of Probabilistic Programming Languages*.

Why Adaptation Matters:

- Vanilla RSA is **static** (fixed priors, fixed lexica).
- Real dialogue is **dynamic**: speakers and listeners adapt over time.

[1] R. D. Hawkins, Frank, and N. D. Goodman, "Characterizing the Dynamics of Learning in Repeated Reference Games".
[2] R. X. D. Hawkins, Frank, and N. D. Goodman, "Convention-formation in iterated reference games".
[3] Schuster and Degen, "I know what you're probably going to say".

Why Adaptation Matters:

- Vanilla RSA is **static** (fixed priors, fixed lexica).
- Real dialogue is **dynamic**: speakers and listeners adapt over time.

Core Idea: Adaptation follows directly from RSA's Bayesian structure:

$$\text{today's posterior} \quad \Rightarrow \quad \text{tomorrow's prior}$$

What gets updated?

- *Lexical meanings* (lexical uncertainty models).
- *Utterance preferences and costs*.

Empirical Evidence:

- Hawkins et al. (2017, 2020): reference game conventions[1,2].
- Schuster & Degen (2020): adapting to uncertainty expressions[3].

Takeaway: RSA naturally captures **learning, personalization, and context sensitivity**.

---

[1] R. D. Hawkins, Frank, and N. D. Goodman, "Characterizing the Dynamics of Learning in Repeated Reference Games".
[2] R. X. D. Hawkins, Frank, and N. D. Goodman, "Convention-formation in iterated reference games".
[3] Schuster and Degen, "I know what you're probably going to say".

Why Study Development?

- RSA models how learners integrate cues: **informativity**, **common ground**, **lexicon**.
- Central question: Do children and adults rely on the **same Bayesian logic**, or does the mechanism itself change?

[1] Bohn, Tessler, and Frank, "Integrating Common Ground and Informativeness in Pragmatic Word Learning"; Bohn, Tessler, Merrick, et al., "Predicting pragmatic cue integration in adults' and children's inferences about novel word meanings."

[2] K. J. Savinelli, Gregory Scontras, and Pearl, "Modeling scope ambiguity resolution as pragmatic inference"; K. Savinelli, Greg Scontras, and Pearl, "Exactly two things to learn from modeling scope ambiguity resolution".

[3] Brochhagen, Franke, and Rooij, "Coevolution of Lexical Meaning and Pragmatic Use"; Carcassi, Schouwstra, and Kirby, "The evolution of adjectival monotonicity".

Language Development: RSA Across the Lifespan

Why Study Development?

- RSA models how learners integrate cues: **informativity**, **common ground**, **lexicon**.
- Central question: Do children and adults rely on the **same Bayesian logic**, or does the mechanism itself change?

Empirical Findings:

- **Word learning:** Age differences reflect **input reliability**, not different reasoning[1].
- **Ambiguity resolution:** RSA explains distributive vs. collective continuity[2].

Cultural Evolution:

- Learners infer semantics from pragmatic use.
- Iterated learning with RSA explains the emergence of **conventions** and **mutual exclusivity** and **diachronic meaning shifts**[3].

Takeaway: RSA unifies perspectives on **developmental trajectories** and **cultural change** in language.

---

[1] Bohn, Tessler, and Frank, "Integrating Common Ground and Informativeness in Pragmatic Word Learning"; Bohn, Tessler, Merrick, et al., "Predicting pragmatic cue integration in adults' and children's inferences about novel word meanings."

[2] K. J. Savinelli, Gregory Scontras, and Pearl, "Modeling scope ambiguity resolution as pragmatic inference"; K. Savinelli, Greg Scontras, and Pearl, "Exactly two things to learn from modeling scope ambiguity resolution".

[3] Brochhagen, Franke, and Rooij, "Coevolution of Lexical Meaning and Pragmatic Use"; Carcassi, Schouwstra, and Kirby, "The evolution of adjectival monotonicity".

Computational-Level Nature:

- RSA is a **computational-level** analysis[1].
- Explains inference problem, not mechanisms of processing.
- Agents (L_1, S_1, L_0) are **idealized rational**.

---

[1] Marr, *Vision*.

[2] Bohn, Tessler, Merrick, et al., "Predicting pragmatic cue integration in adults' and children's inferences about novel word meanings."

[3] Franke and Degen, "Reasoning in Reference Games".

[4] Griffiths, Lieder, and N. D. Goodman, "Rational Use of Cognitive Resources".

[5] Cohn-Gordon, N. Goodman, and Potts, "Pragmatically Informative Image Captioning with Character-Level Inference".

Levels of Analysis: What RSA Explains (and Doesn't)

Computational-Level Nature:

- RSA is a **computational-level** analysis[1].
- Explains inference problem, not mechanisms of processing.
- Agents (L_1, S_1, L_0) are **idealized rational**.

Where Resource Limits Enter:

- **Optimality parameter α:** lower $\alpha \Rightarrow$ weaker utility maximization; may vary by age/demographics[2].
- **Recursion depth:** limited layers of reasoning[3].
- **Sampling approximations:** RSA as resource-bounded rationality[4].

---

[1] Marr, *Vision*.

[2] Bohn, Tessler, Merrick, et al., "Predicting pragmatic cue integration in adults' and children's inferences about novel word meanings."

[3] Franke and Degen, "Reasoning in Reference Games".

[4] Griffiths, Lieder, and N. D. Goodman, "Rational Use of Cognitive Resources".

[5] Cohn-Gordon, N. Goodman, and Potts, "Pragmatically Informative Image Captioning with Character-Level Inference".

Levels of Analysis: What RSA Explains (and Doesn't)

Computational-Level Nature:

- RSA is a **computational-level** analysis[1].
- Explains inference problem, not mechanisms of processing.
- Agents (L_1, S_1, L_0) are **idealized rational**.

Where Resource Limits Enter:

- **Optimality parameter α:** lower $\alpha \Rightarrow$ weaker utility maximization; may vary by age/demographics[2].
- **Recursion depth:** limited layers of reasoning[3].
- **Sampling approximations:** RSA as resource-bounded rationality[4].

Incrementality:

- Human comprehension is **incremental**, not utterance-final.
- Incremental RSA: reasoning over prefixes/continuations[5].

---

[1] Marr, *Vision*.

[2] Bohn, Tessler, Merrick, et al., "Predicting pragmatic cue integration in adults' and children's inferences about novel word meanings."

[3] Franke and Degen, "Reasoning in Reference Games".

[4] Griffiths, Lieder, and N. D. Goodman, "Rational Use of Cognitive Resources".

[5] Cohn-Gordon, N. Goodman, and Potts, "Pragmatically Informative Image Captioning with Character-Level Inference".

Why RSA for NLP?:

- Literal models often fail to capture **contextual goals** and **pragmatic meaning**.
- RSA provides a **probabilistic, goal-sensitive layer** for interpretation and generation.

Two Main Applications:

1. **Learning semantics through pragmatic reasoning**
   - Train models assuming data come from **pragmatic speakers**.
   - Improves semantic representations[1][2].
2. **Pragmatic wrapper around neural models**
   - Combine RSA reasoning with pretrained systems for: Image captioning[3], Vision-language navigation[4], Summarization & MT[5].
   - Boosts **informativeness**, **contrastiveness**, and **context awareness**.

Key Insight: RSA bridges **linguistic theory** and **AI systems**, enabling models that reason about **speaker goals**, **listener expectations**, and **contextual constraints**.

[1] Monroe and Potts, *Learning in the Rational Speech Acts Model*.
[2] Monroe, R. X. Hawkins, et al., "Colors in Context".
[3] Cohn-Gordon, N. Goodman, and Potts, "Pragmatically Informative Image Captioning with Character-Level Inference".
[4] Fried et al., "Speaker-follower models for vision-and-language navigation".
[5] Shen et al., "Pragmatically Informative Text Generation".

Table of Contents

Conclusion

- RSA models formalize language understanding as **probabilistic, recursive social reasoning**.
- They provide a computational implementation of **Gricean pragmatics**.
- Extensions of RSA capture:
  - Lexical and semantic uncertainty
  - Non-literal language via QUD inference
  - Contextual and epistemic variability
  - Social goals through complex utility functions
- Modeling involves both **mathematical abstraction** and **probabilistic programming**.
- Practical modeling requires careful design of:
  - World states, utterances, priors
  - Linking functions to behavioral data
  - Parameter estimation and model comparison
- RSA is being actively extended to:
  - Compositional semantics
  - Adaptation and learning
  - Language development
  - NLP applications

[1] **Gregory Scontras, Tessler, and Franke,** ***A practical introduction to the Rational Speech Act modeling framework*****.**

- THE END -

Thank you for your attention

Acknowledgements
Special thanks to Prof. Tan Zhi Xuan for guidance and support throughout this presentation.

Contact
nqduc@u.nus.edu