Linguistics: Dialogue, Conversation, and Interaction

EECS 183/283a: Natural Language Processing

Today

- Language use in context (pragmatics)
- Multi-agent foundations of interaction
- Interaction dynamics
- Why do we care about language-based interaction?

Pragmatics: Evaluation of λ-expressions in Context

What color is Pepper?

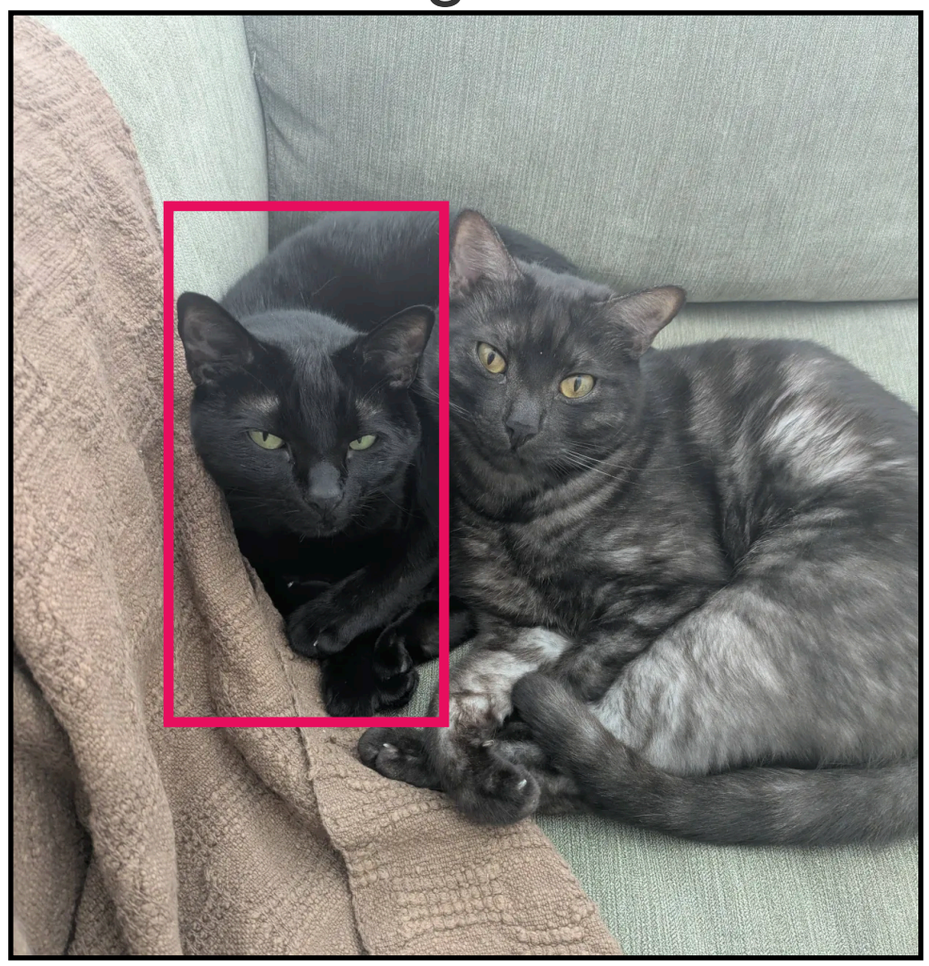
λ f . (Color(f) ∧ f(Pepper))

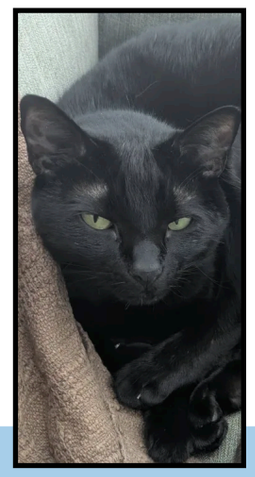
looking for a function

that looks up the color of its argument

and is true for the `Pepper` primitive

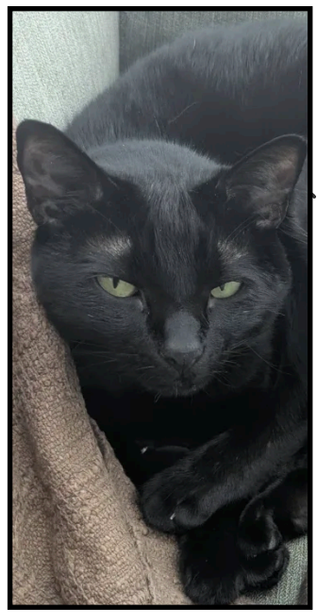
Image i



$$[\![\text{Pepper}]\!]^i = \{$$  $$\}$$

Pragmatics: Evaluation of λ-expressions in Context

What color is Pepper?

`λ f . (Color(f) ∧ f(`  `)`

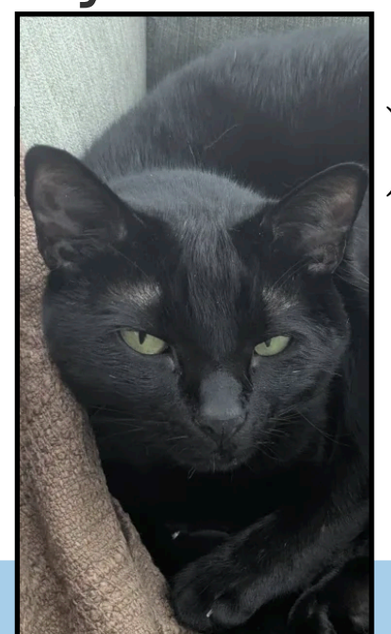
looking for a function

that looks up the color of its argument

functions of type $\langle p \rightarrow t \rangle$

ID	Lexeme	implementation
1	small	`λ p . p.size == small`
2	purple	`λ p . p.color == purple`
…	…	…
N	black	`λ p . p.color == black`

Iterate over the functions that check the `color` property of an entity

$$[\![\text{purple}(\text{Pepper})]\!]^i = \text{purple}($$  $$)$$

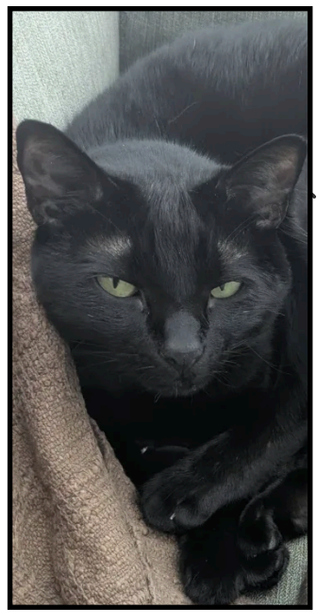
Pragmatics: Evaluation of λ-expressions in Context

What color is Pepper?

λ f . (Color(f) $\wedge$ f( ))

looking for a function

that looks up the color of its argument



functions of type $\langle p \rightarrow t \rangle$

ID	Lexeme	implementation
1	small	λ p . p.size == small
2	purple	λ p . p.color == purple
…	…	…
N	black	λ p . p.color == black

Iterate over the functions that check the `color` property of an entity

$$[\![\text{purple}(\text{Pepper})]\!]^i = \text{purple}([\![\text{Pepper}]\!]^i) = \text{False}$$

$$[\![\text{black}(\text{Pepper})]\!]^i = \text{black}([\![\text{Pepper}]\!]^i) = \text{True}$$

$$[\![\lambda f.(\text{Color}(f) \wedge f(\text{Pepper}))]\!]^i = \{\text{black}\}$$

Semantics vs. Pragmatics

What color is Pepper?

$\lambda f.(\text{Color}(f) \wedge f(\text{Pepper}))$

$[\![\lambda f.(\text{Color}(f) \wedge f(\text{Pepper}))]\!]^i$

$\{\text{black}\}$

Semantics: mapping from surface form (sequence of tokens) to formal executable representation

Pragmatics: executing the logical form against some context to acquire its denotation

Speech Acts

- Our interpretation of utterances used in context often goes beyond the literal (formal) meaning

- By interpreting speech as action, we can ascribe intent to utterances that isn’t obvious from their formal representation

Do you mind if I sit next to you?

Yeah (go ahead)

No (I don’t mind)

Sorry, someone is coming

Examples from Linda Tjuatja

Presupposition

- Propositions that must be true about a world in order to compute the denotation of a particular sentence

- In other words: implicit assumptions made by utterances

Pepper owns a house — Can be true or false

Pepper's house is big — Computing the truth value requires that, in our world, there is an entity `x` such that `house(x)` and `owns(Pepper, x)`

Awareness of presupposition in speech is very useful for critical analysis of persuasive speech, e.g. in politics

Implicature

- Propositions suggested by an utterance, but not explicitly expressed

- Meaning is determined by reasoning about alternatives

Do you know what they weather will be like today?

You should bring your umbrella.

Context: in San Francisco

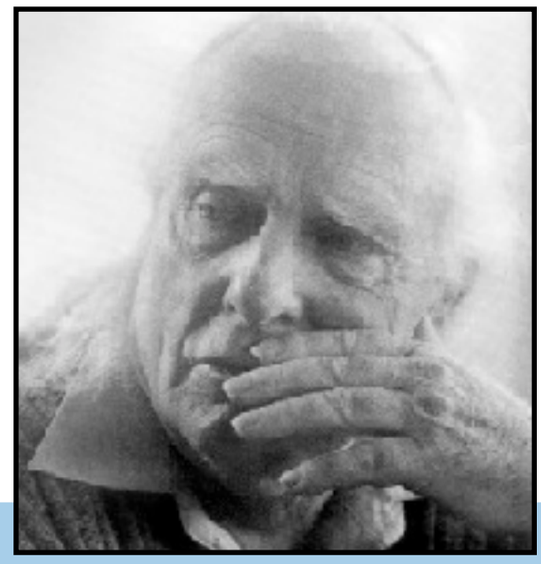
↝ Heavy rain is possible

Alternatives:

- *no*
- *yes*
- *low 60s*
- *chance of precipitation is 30%*

Principles behind Implicature: Gricean Maxims

- General principles we believe we mutually hold about how what kinds of utterances we should add to conversation given what's been said so far:
  - **Quantity:** utterances should contain just the right amount of information — not too little or too much
  - **Truth:** utterances should not contain falsehoods
  - **Relation:** utterances should be relevant to what's been said before
  - **Manner:** utterance form and meaning should be clear
- What happens when we contribute utterances that break these principles?



Grice 1957, Logic and conversation

Principles behind Implicature: Gricean Maxims

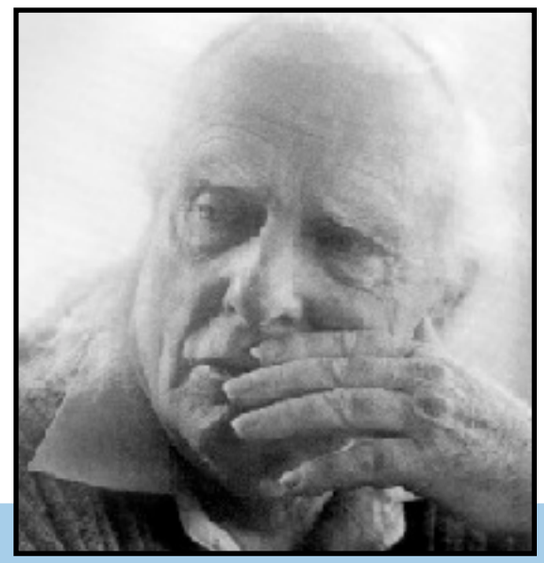
- **Flouting** conversation maxims is “breaking” them under the assumption the listener knows the speaker is intentionally breaking them
  - E.g., flouting relevance:

Do you know what they weather will be like today?

You should bring your umbrella.

- **Violating** maxims is breaking them under the assumption the listener *won’t* believe a maxim has been broken

My dog ate my homework.



Grice 1957, Logic and conversation

Common Ground

During interaction, we maintain some representation of what we believe is mutually known by conversation participants

- Mutually known: I know it, I believe you know it, I believe you believe I know it, I believe you believe I believe you know it, …

- What can be in the common ground?
  - Principles guiding how we interact with one another (e.g., Gricean maxims, shared lexicon)
  - Propositions about the world, values and beliefs
  - Things in our shared environment, including things we are paying attention to

Common Ground

- This allows us to reason about what is not in the common ground
  - E.g., facts, beliefs, etc. that we believe the other does not know

- By maintaining models of others' beliefs, we can reason about how they might interpret our utterances

- If we want to bridge a belief gap between ourselves and another conversation participant, we can rely on the the rules governing language use and interpretation that (we believe) are in the common ground

Multi-Agent Foundations of Interaction

- You’re playing Minecraft, and want to build a house
  - Your actions influence the state of the world
  - But there’s also stochasticity (e.g., animals, enemies, villagers appear and might destroy your house, but their behavior is (relatively) predictable)

Multi-Agent Foundations of Interaction

- You’re playing Minecraft, and want to build a house
  - Your actions influence the state of the world
  - But there’s also stochasticity (e.g., animals, enemies, villagers appear and might destroy your house, but their behavior is predictable)
- But what happens when another player comes to the game?
  - They also want to build a house
  - Maybe they will take some of your resources from you
  - Their behavior is not completely stochastic, though!

Multi-Agent Foundations of Interaction

To best model another player, we may want to keep track of their:

- **Beliefs:** what information are they using to make decisions? How do they perceive the world, and how do they build an internal model of the world as they act in it?
- **Goals:** what do they want to get done? Do they share the same goals, are their goals orthogonal, or are they trying to sabotage mine?
- **Intentions:** how will they attempt to execute their goals? What skills do they have and what strategies are they likely to take?
- **Model of *me*:** if it's useful for me to reason about them, they are probably reasoning about me, too — how does this influence their actions?

Cooperative Interaction

- Let’s assume my partner and I share the same (high-level) goal, and it’s in the common ground
  - Goal: build a house

- It’s likely optimal for me to choose actions in a way that depends on my partner’s actions, to avoid redundancy and execute the goals more efficiency
  - I’ll gather wood while my partner places it in the right spots to create a foundation

- Some environments might require that I and my partner take different actions at the same time
  - E.g., pressing paired switches

Cooperative Interaction

- How can we more successfully coordinate with one another, especially under uncertainty over how the environment works?

- How can I better model my partner?
  - What they observe and know?
  - What they are trying to do?
  - What are their plans to act?

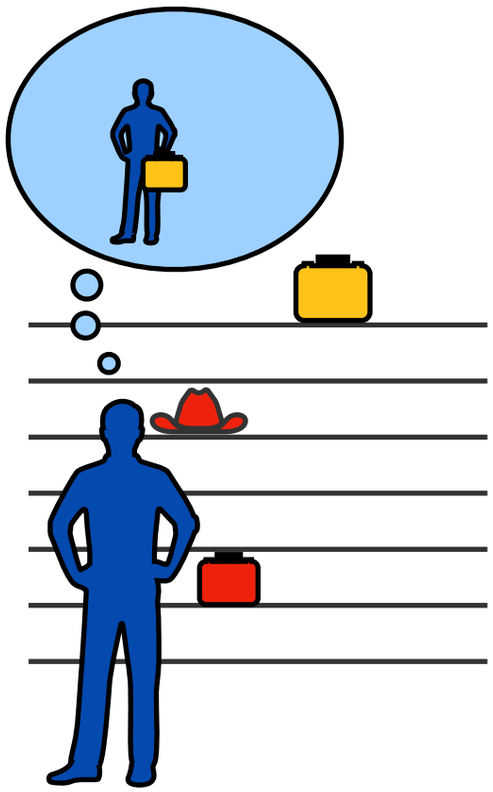
- How can I influence my partner?
  - By sharing information with them?
  - By telling them what to do?
  - By teaching them about how to act?

Language

If we have language in common ground, then I can:

- Share information with my partner via statements
- Ask questions of my partner
- Teach my partner something
- Learn something from my partner
- Give my partner instructions
- Coordinate on what we should do towards our shared goals

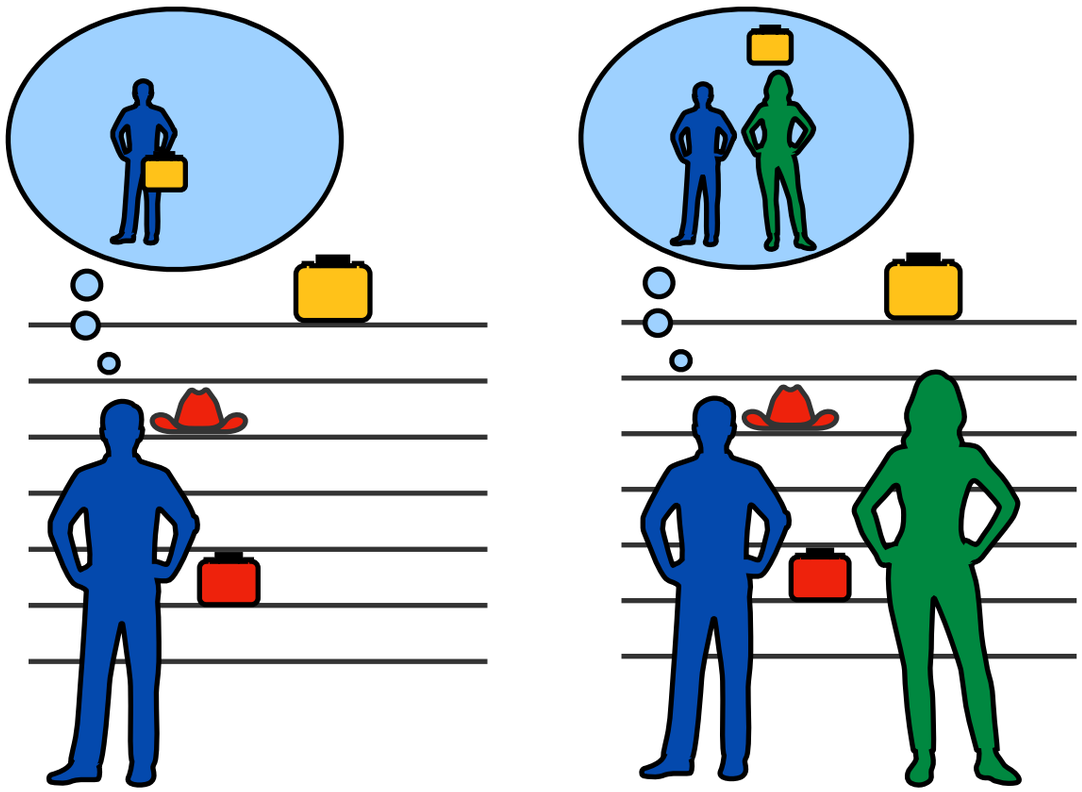
Language Use



<u>My goal:</u>
Get the suitcase

but… I'm too short

Language Use



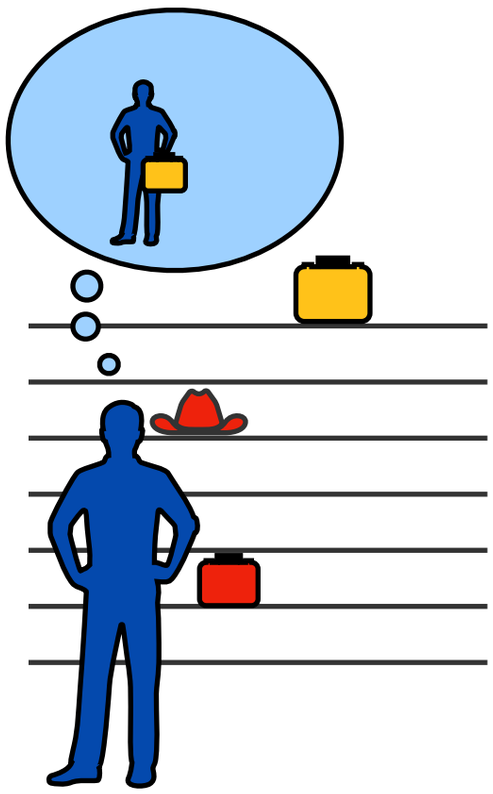
My goal:
Get the suitcase

but… I'm too short

Observation:
Green person can reach it

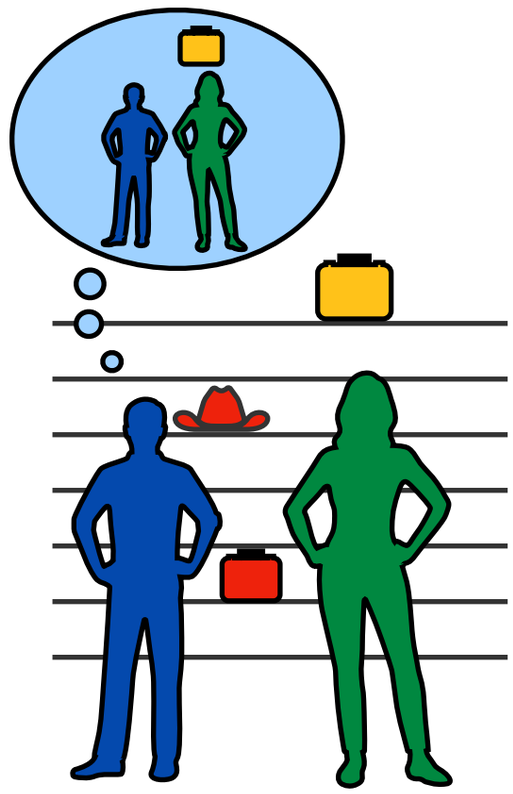
Optimal action:
Green person gives it to me

Language Use



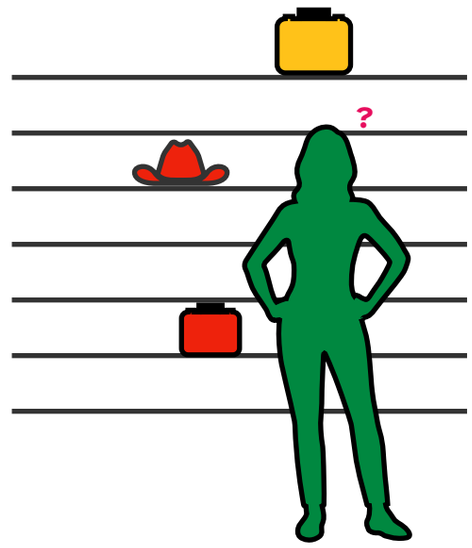
My goal:
Get the suitcase

but… I'm too short



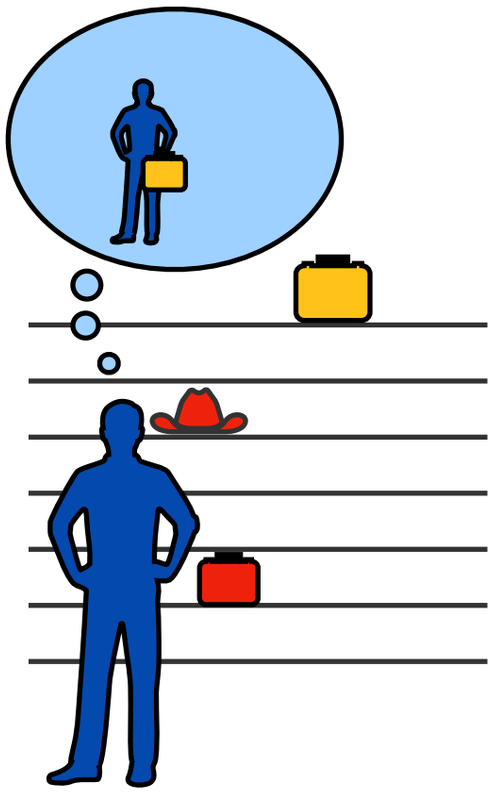
Observation:
Green person can reach it

Optimal action:
Green person gives it to me



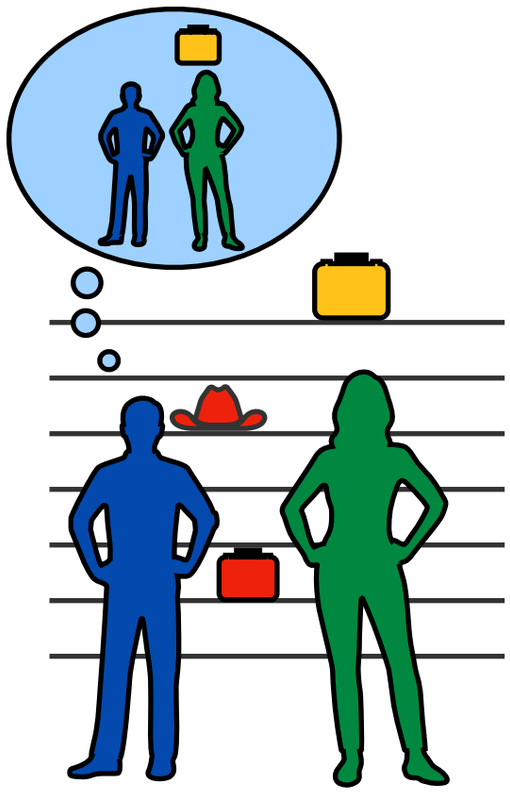
Belief:
Green person doesn't know my goal

Language Use



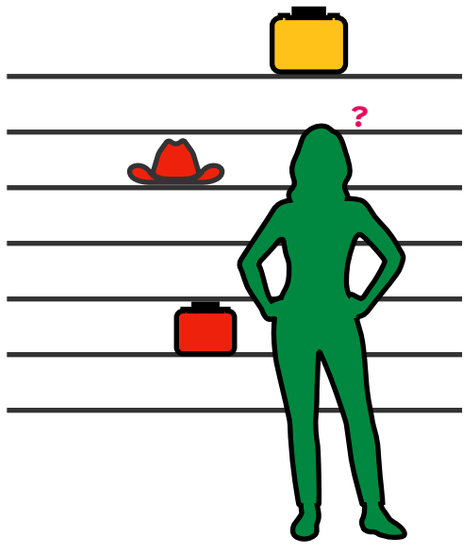
My goal:
Get the suitcase

but… I'm too short



Observation:
Green person can reach it

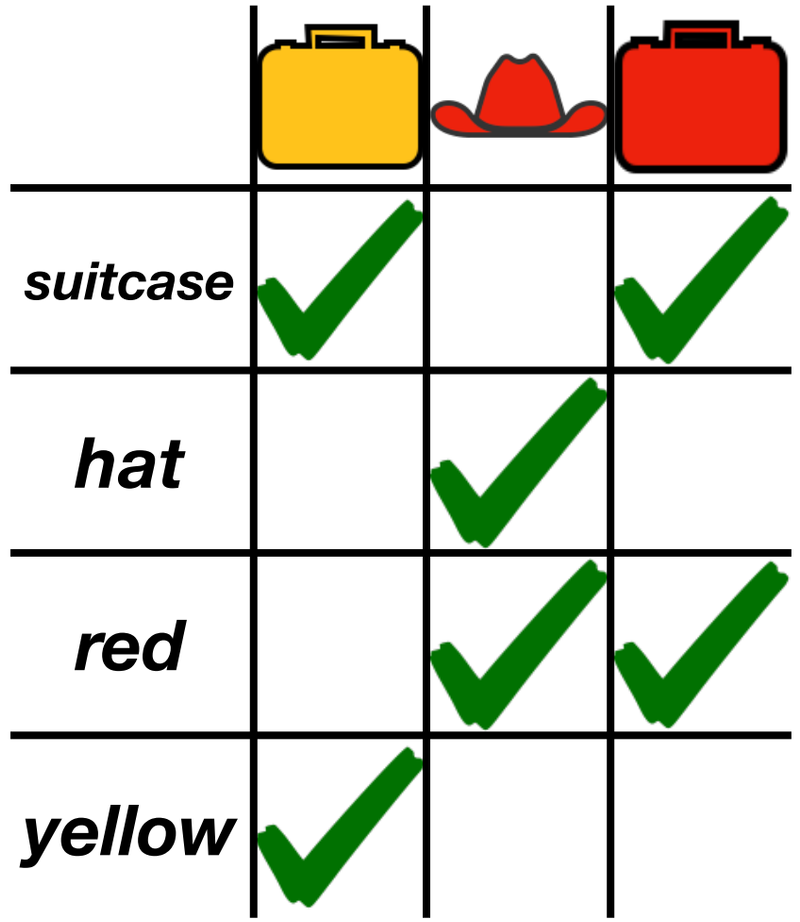
Optimal action:
Green person gives it to me



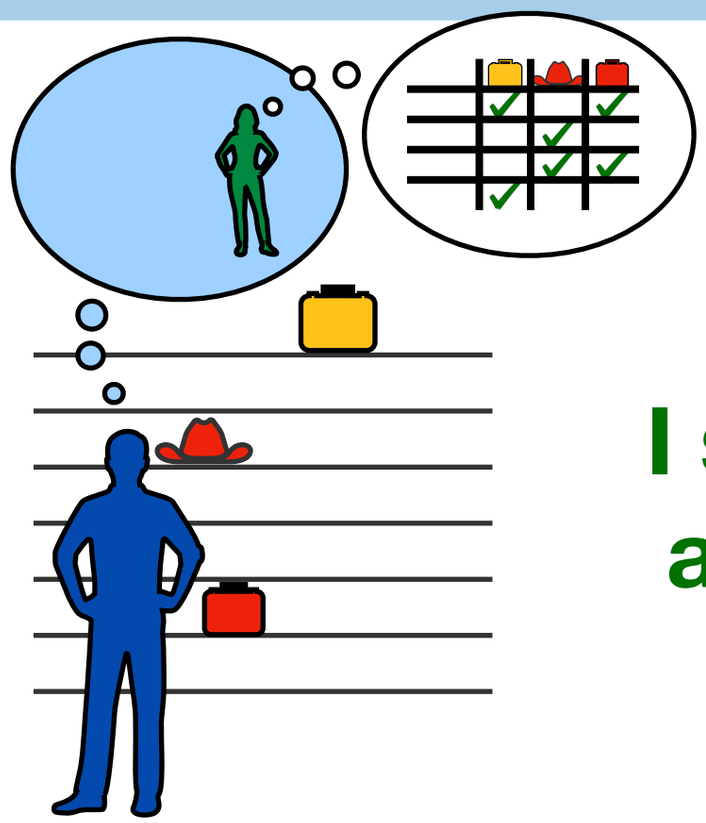
Belief:
Green person doesn't know my goal

How can I get the suitcase?

Theory of mind reasoning: I share a *lexicon* with the green person



Language Use



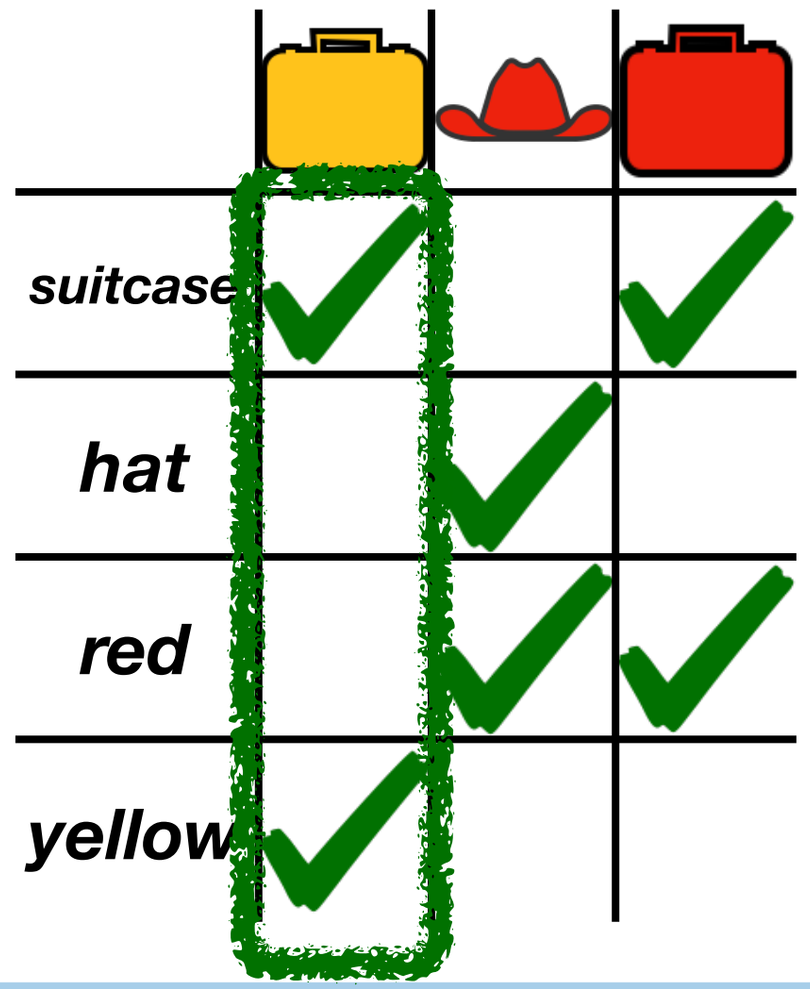
My goal:
Get the suitcase

but… I'm too short

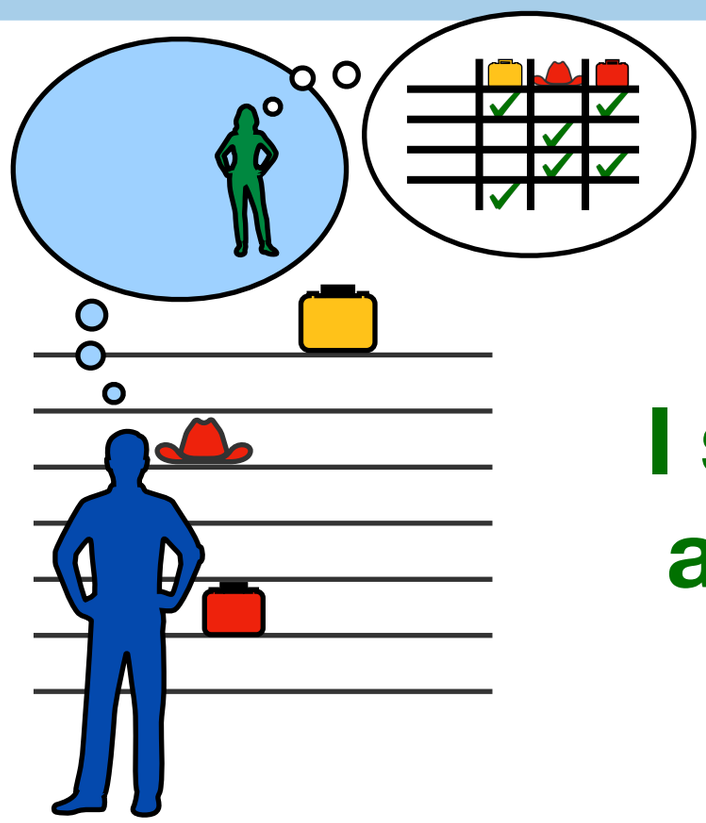
I should choose an action from which my listener can infer my goal

How can I get the suitcase?

Theory of mind reasoning: I share a *lexicon* with the green person



Language Use



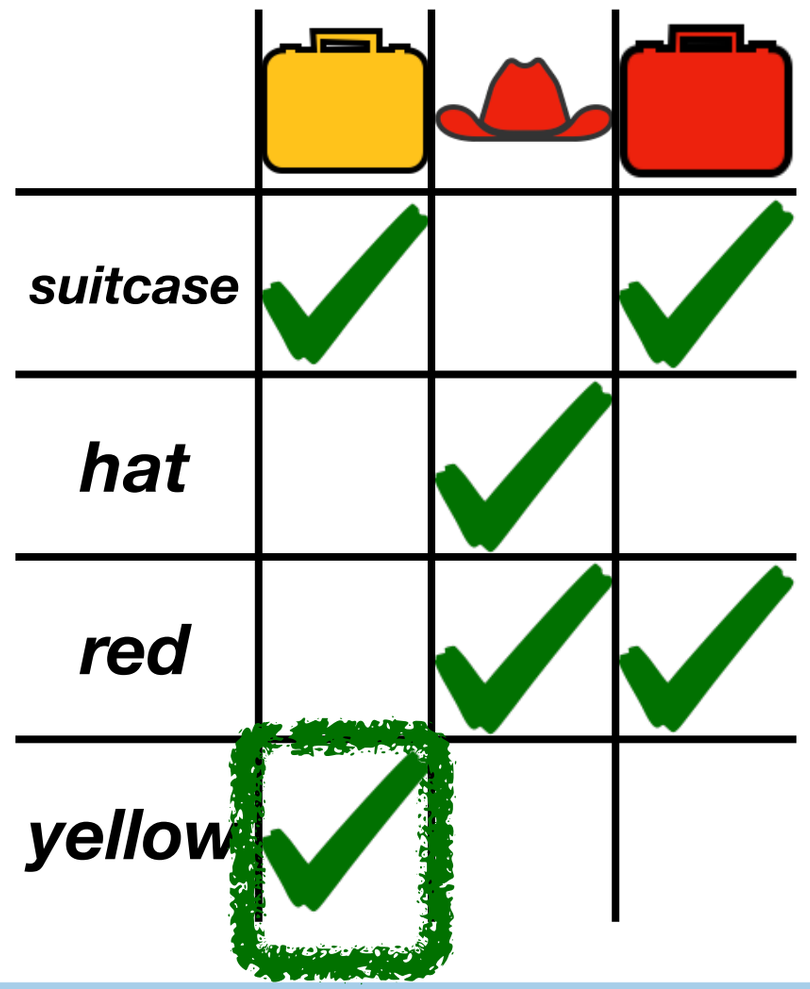
My goal:
Get the suitcase

but… I'm too short

I should choose an action from which my listener can infer my goal uniquely.

How can I get the suitcase?

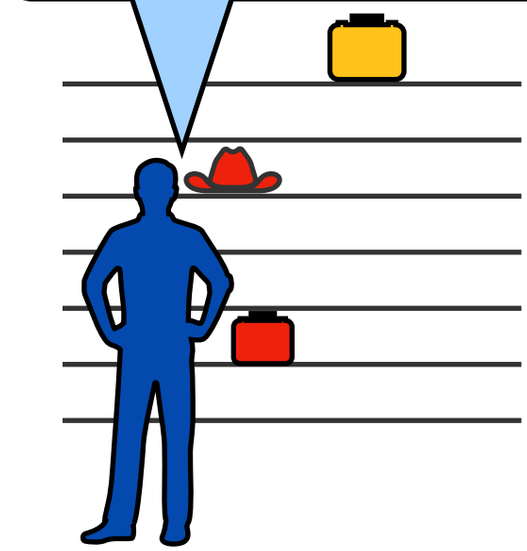
Theory of mind reasoning: I share a *lexicon* with the green person



RSA; Frank and Goodman 2012

Language Use

please hand me the yellow suitcase



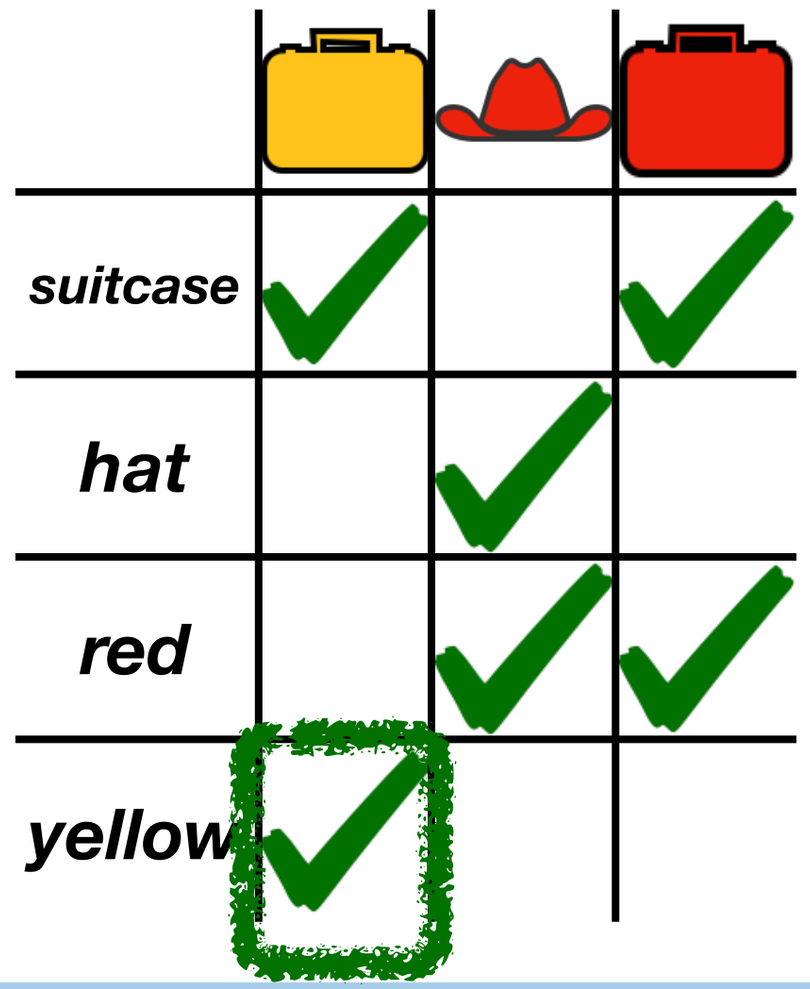
My goal:
Get the suitcase

but… I'm too short

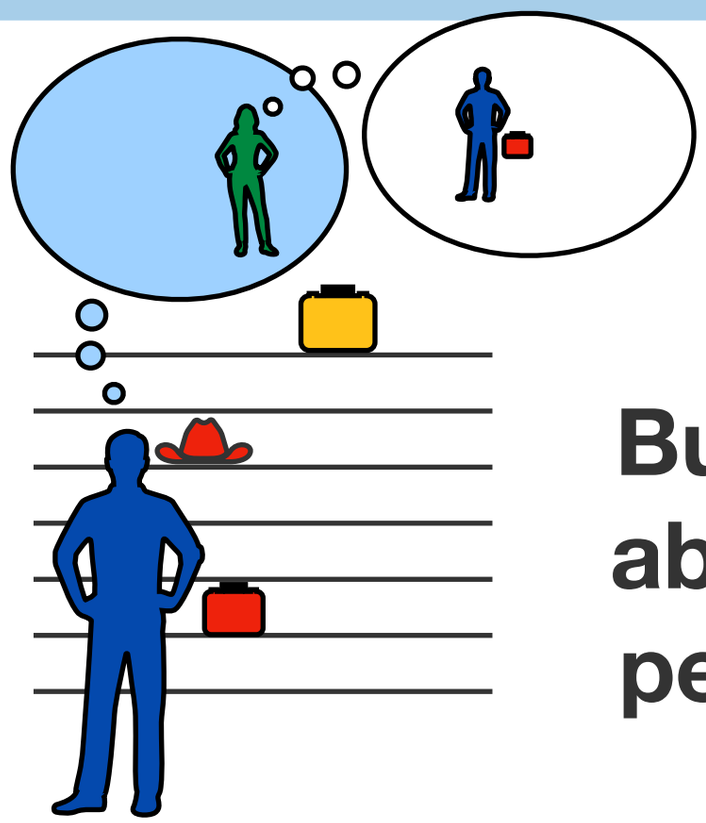
I should choose an action from which my listener can infer my goal uniquely.

How can I get the suitcase?

Theory of mind reasoning: I share a *lexicon* with the green person



Language Use



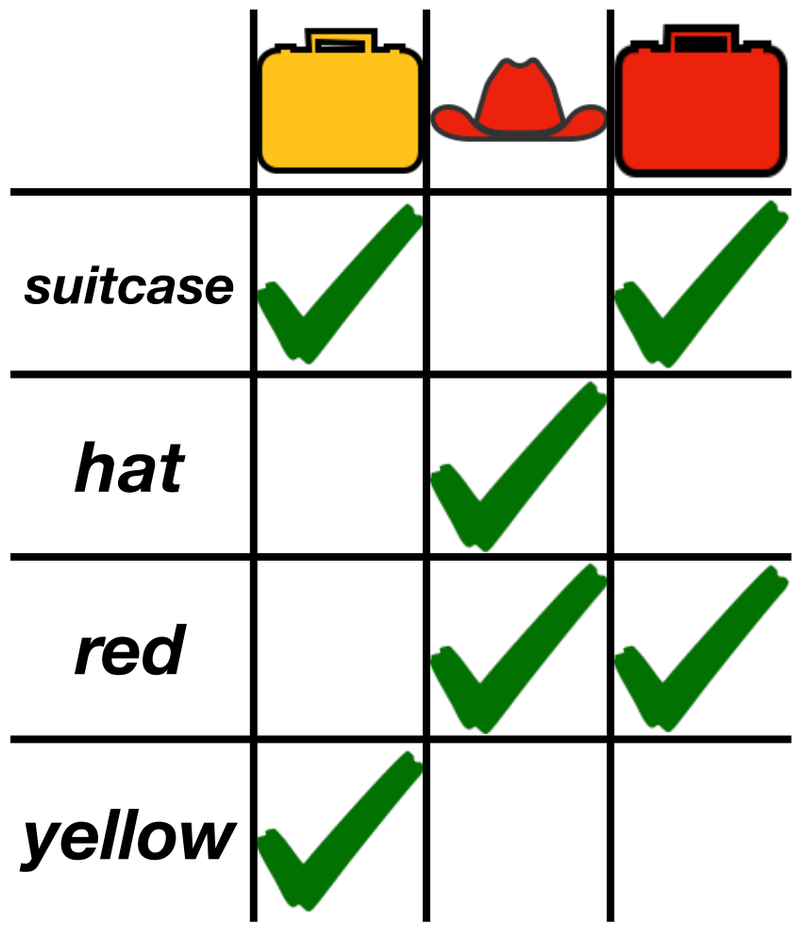
My goal:
Get the suitcase

but… I'm too short

But I can also reason about what the green person might believe about me.

How can I get the suitcase?

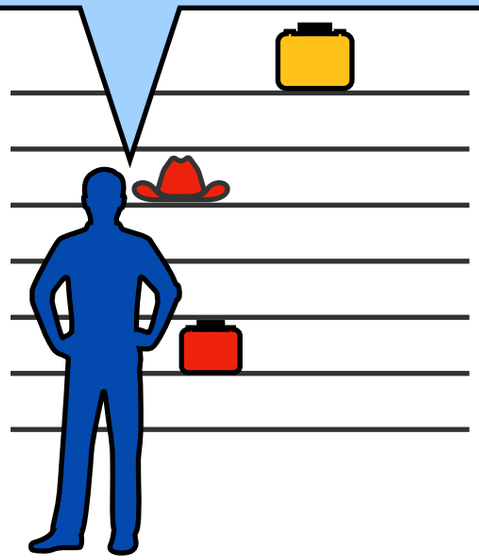
Theory of mind reasoning: I share a *lexicon* with the green person



RSA; Frank and Goodman 2012

Language Use

please hand me the ~~yellow~~ suitcase



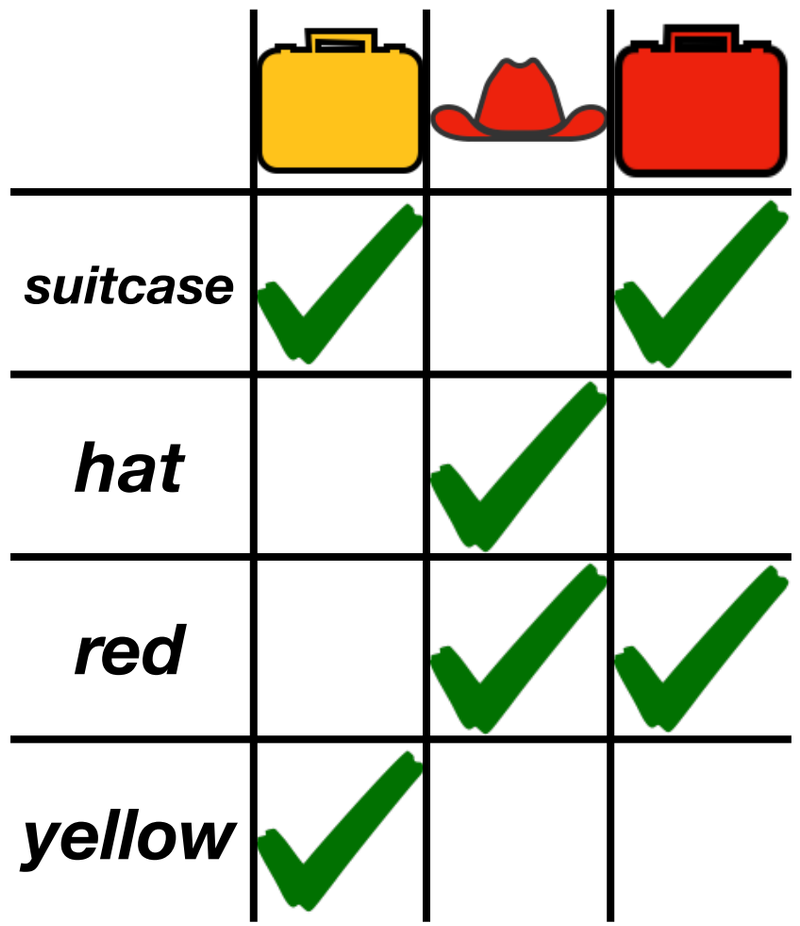
My goal:
Get the suitcase

but… I'm too short

But I can also reason about what the green person might believe about me.

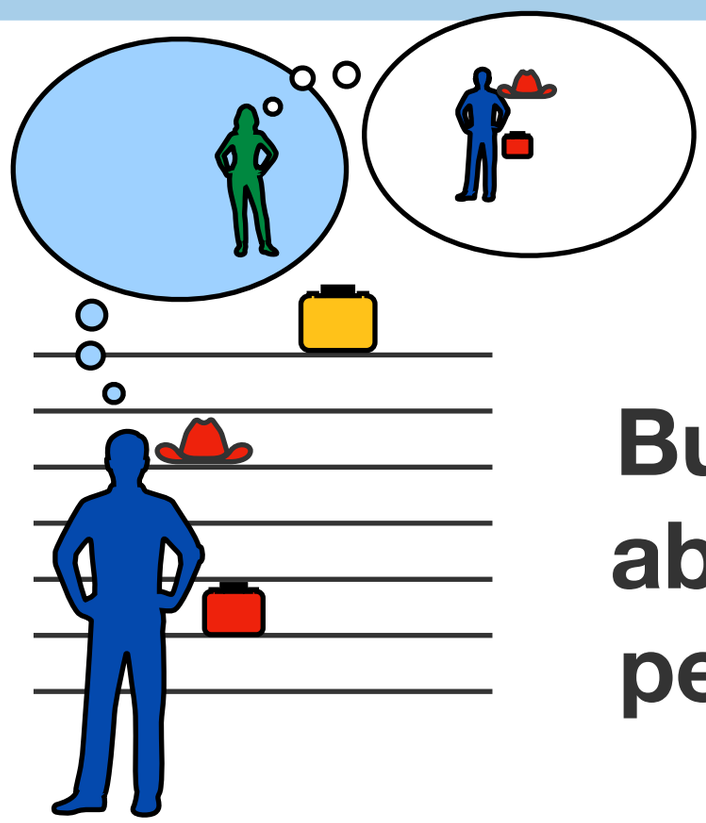
How can I get the suitcase?

Theory of mind reasoning: I share a *lexicon* with the green person



RSA; Frank and Goodman 2012

Language Use



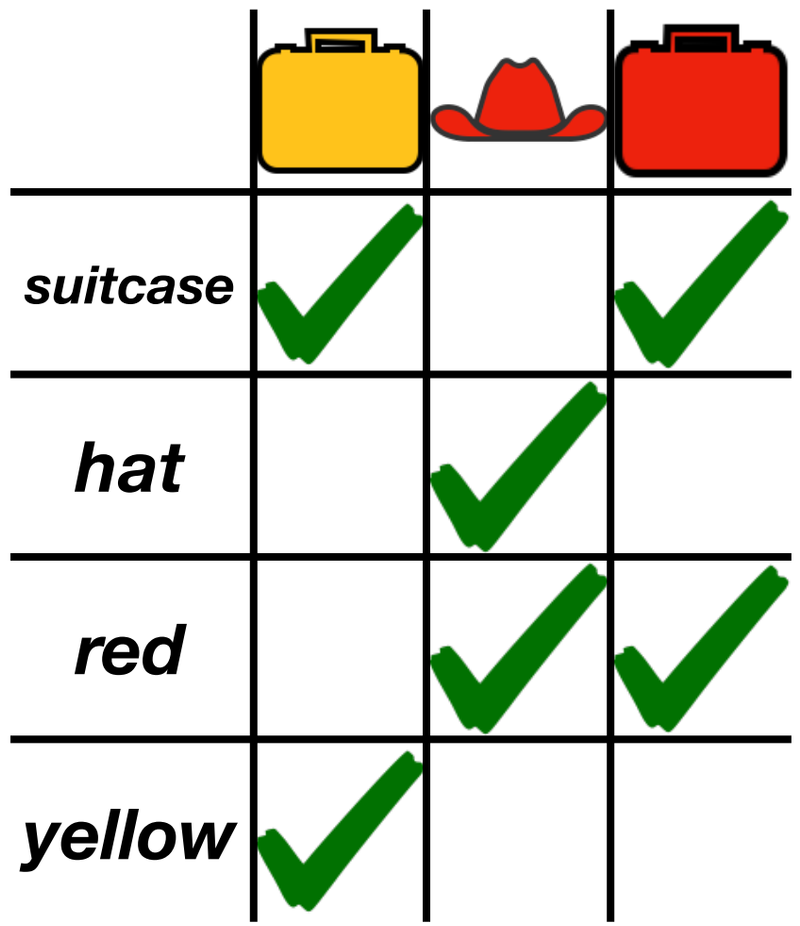
My goal:
Get the suitcase

but… I'm too short

But I can also reason about what the green person might believe about me.

How can I get the suitcase?

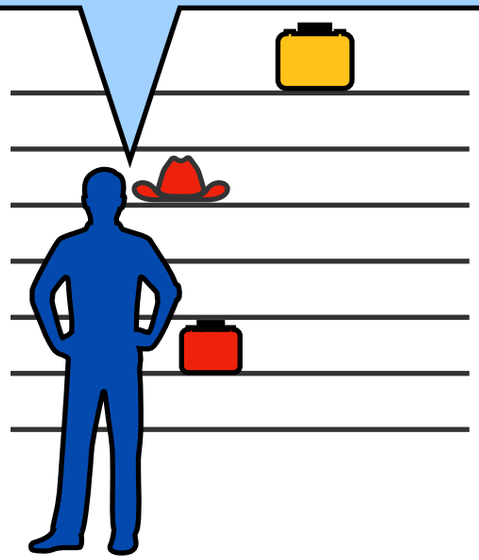
Theory of mind reasoning: I share a *lexicon* with the green person



RSA; Frank and Goodman 2012

Language Use

please hand me the ~~yellow suitcase~~



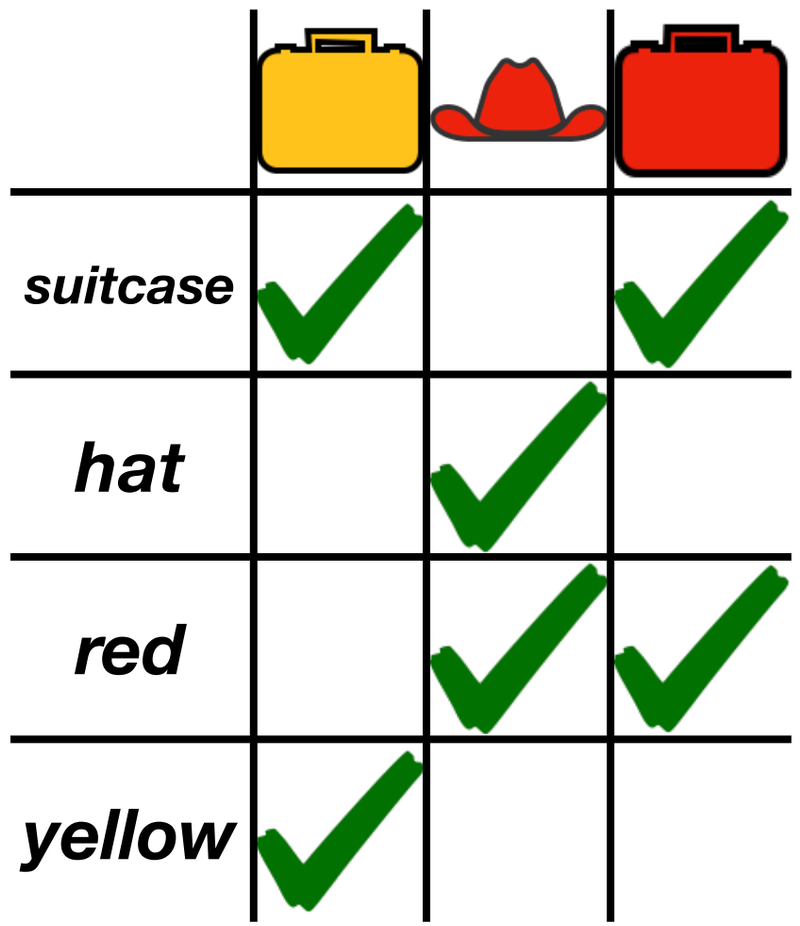
My goal:
Get the suitcase

but… I'm too short

But I can also reason about what the green person might believe about me.

How can I get the suitcase?

Theory of mind reasoning: I share a *lexicon* with the green person



Language in Multi-Agent Environments

- Observation space now includes utterances made by other agent(s)

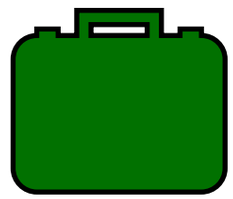
- Action space now includes the ability to produce an utterance

Interaction Dynamics

- At the beginning of an interaction, we might have significant uncertainty over other agents
  - Over their goals, beliefs, and skills
  - And also over how they use language
- But over time, we converge to more similar representations
  - By building models of one another from observing their behavior
  - By explicitly resolving uncertainties via language use
  - This refines our expectations of other agents
  - Certainty in our expectations allows us to take communicative shortcuts

<u>hand me</u> the yellow suitcase

hand me the green suitcase



the blue one

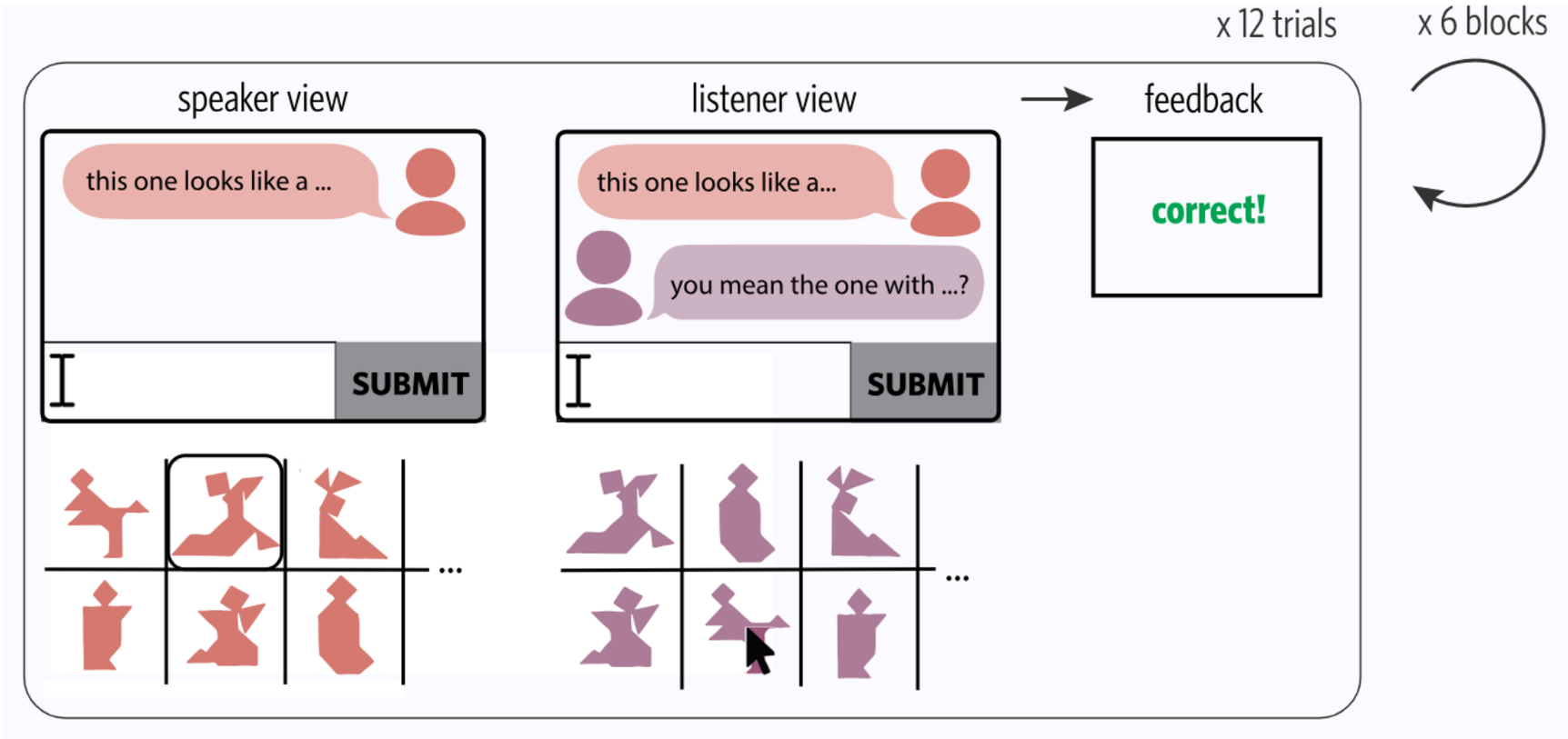
red

Pickering and Garrod 2004

Interaction Dynamics

The influence of learning on interaction dynamics:

- Agent policies adapt as they learn about other agents
- This influences the observations other agents make of. them
- Which in turn influences how other agents adapt
- E.g., convention formation in iterated reference games

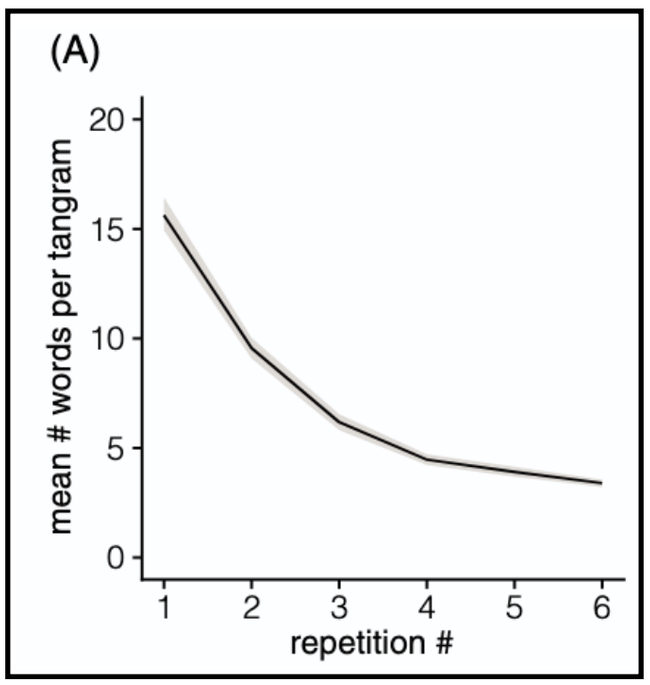


Hawkins et al. 2020

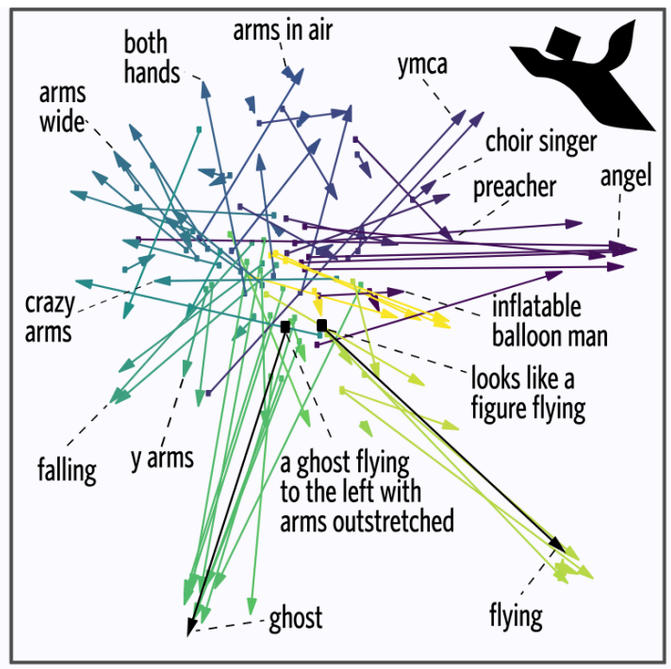
Interaction Dynamics

The influence of learning on interaction dynamics:

- Agent policies adapt as they learn about other agents
- This influences the observations other agents make of. them
- Which in turn influences how other agents adapt
- E.g., convention formation in iterated reference games



References get shorter…



but their exact forms are dependent on the interaction history

Hawkins et al. 2020

Interaction Dynamics

- **Challenges for learning in multi-agent environments:**
  - Agents need to learn their first-order policy (goal, observation → action)
  - But they also might need to model how their adaptations influence the behavior of other agents, including the introduction of new shared abstractions (i.e., words and conventions)!
- **Challenges for evaluating multi-agent systems:**
  - Dynamics depend heavily on initial conditions: (uncertainty over) variation across agent partners' beliefs, goals, and intentions
  - E.g., in teaching contexts, a teacher will adapt their language and pedagogical strategy to the learner's existing knowledge and skills

Interaction Dynamics

What's a good place to get gluten free pastries in Berkeley?

For dedicated gluten-free pastries in Berkeley, your absolute best bet is **Mariposa Baking Company.**

A gluten-free pastry lover in Berkeley! You're in luck because Berkeley is a haven for GF enthusiasts! Here are some top recommendations for gluten-free pastries in Berkeley:

1. Bette's Oceanview Diner (1807 4th St, Berkeley) — This beloved diner has

...

Interaction Dynamics

What's a good place to get gluten free pastries in Berkeley?

For dedicated gluten-free pastries in Berkeley, your absolute best bet is **Mariposa Baking Company.**

That's too far from me. Do you have a recommendation near downtown Berkeley?

A gluten-free pastry lover in Berkeley! You're in luck because Berkeley is a haven for GF enthusiasts! Here are some top recommendations for gluten-free pastries in Berkeley:

1. Bette's Oceanview Diner (1807 4th St, Berkeley) — This beloved diner has ...

Interaction Dynamics

What's a good place to get gluten free pastries in Berkeley?

For dedicated gluten-free pastries in Berkeley, your absolute best bet is **Mariposa Baking Company.**

A gluten-free pastry lover in Berkeley! You're in luck because Berkeley is a haven for GF enthusiasts! Here are some top recommendations for gluten-free pastries in Berkeley:

1. Bette's Oceanview Diner (1807 4th St, Berkeley) — This beloved diner has ...

Are any of these open on Sundays?

Interaction Dynamics

What's a good place to get gluten free pastries in Berkeley?

For dedicated gluten-free pastries in Berkeley, your absolute best bet is **Mariposa Baking Company.**

That's too far from me. Do you have a recommendation near downtown Berkeley?

A gluten-free pastry lover in Berkeley! You're in luck because Berkeley is a haven for GF enthusiasts! Here are some top recommendations for gluten-free pastries in Berkeley:

1. Bette's Oceanview Diner (1807 4th St, Berkeley) — This beloved diner has ...

Are any of these open on Sundays?

User behavior changes depending on system behavior!

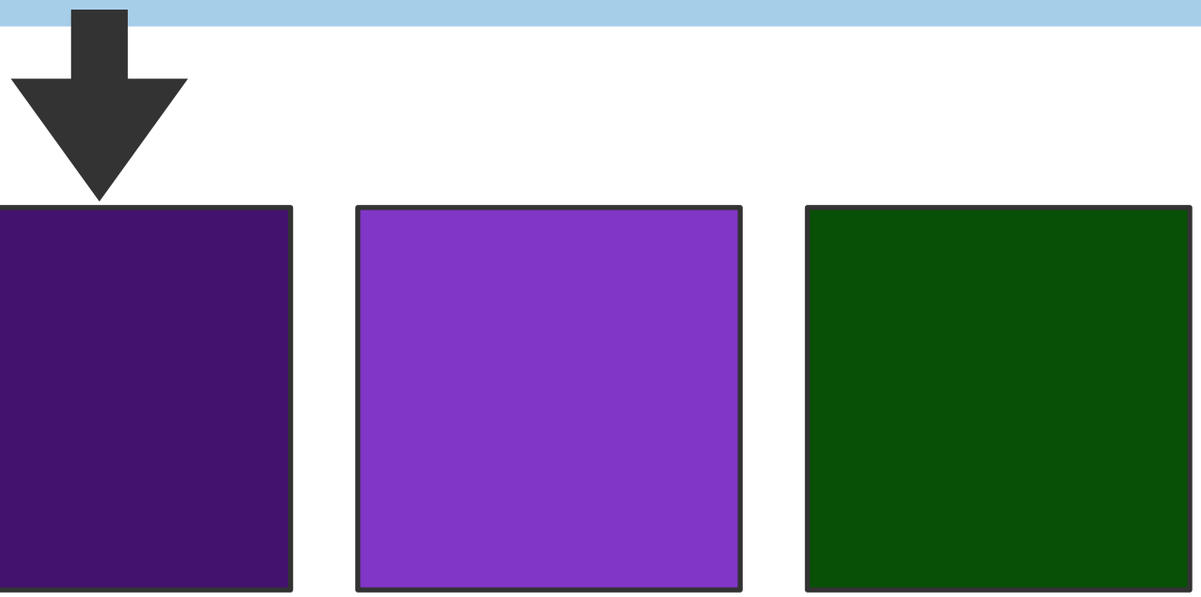
Elements of Scenario Design

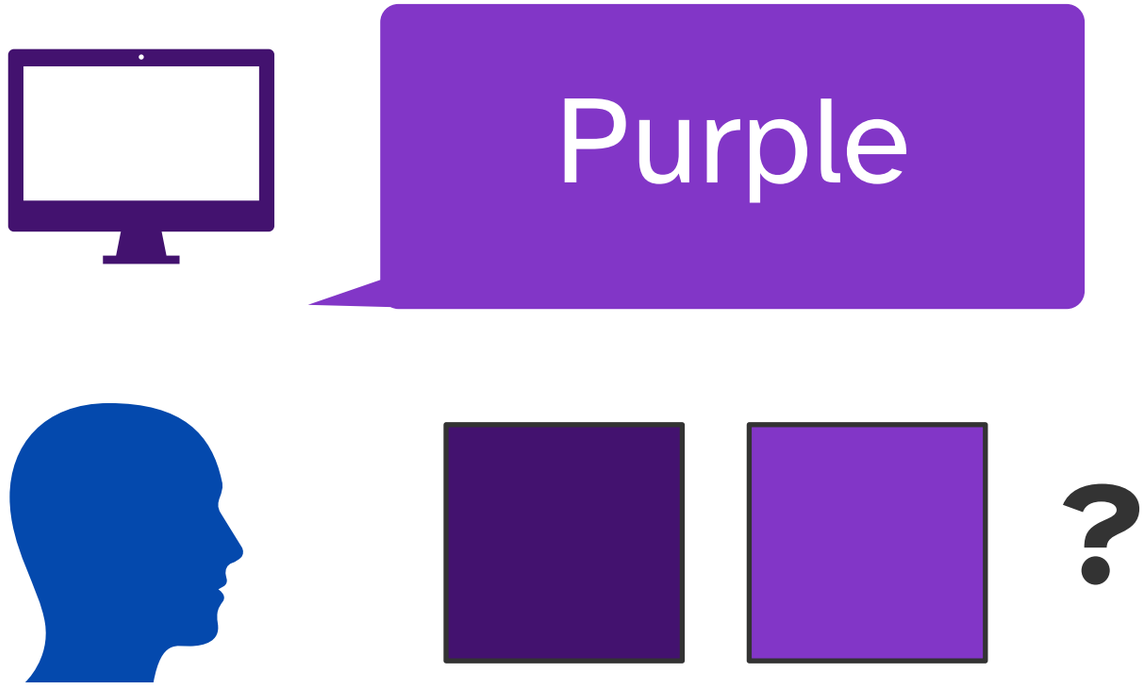
- Interaction dynamics also depend heavily on the properties of the context itself:
  - Incentive structure
  - Environment design — perception and action, novelty
  - Participants — how many, any existing structures among them, roles, a priori asymmetries
  - Communication channel
- Work in computational linguistics, psycholinguistics, and cognitive science aims to characterize the relationship between scenario design and linguistic behavior

Simple Multi-Agent Scenario: Reference Game

- **Environment:** set of candidate referents, available to both players
- **Players**
  - Speaker: knows the identity of a target object
  - Listener: no privileged information
- **Shared goal:** listener picks out the target object
- **Actions**:
  - Speaker: natural language
  - Listener: selection of candidate referent
- **Communication channel:** unidirectional, single utterance (no dialogue)

Simple Multi-Agent Scenario: Reference Game





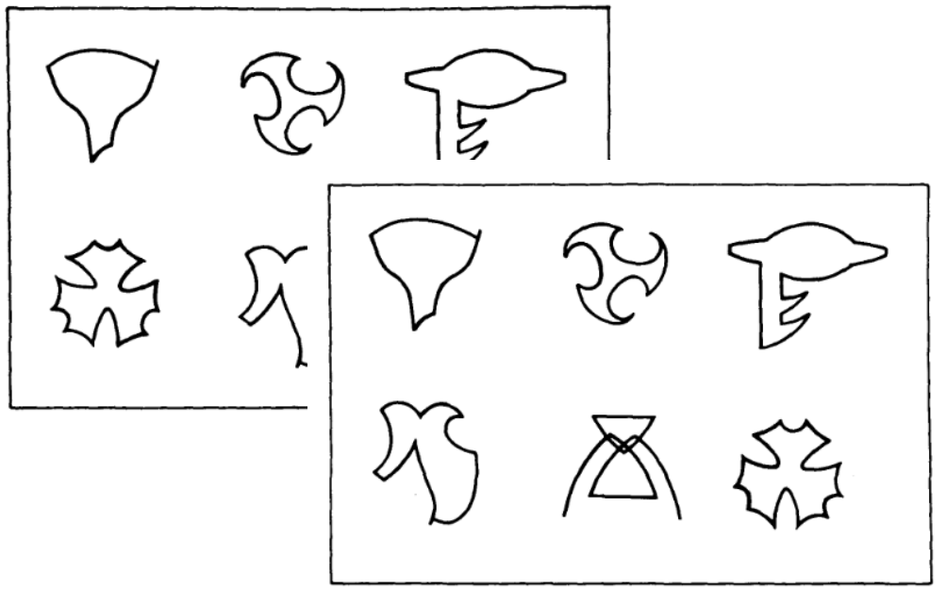
Simple Multi-Agent Scenario: Reference Game

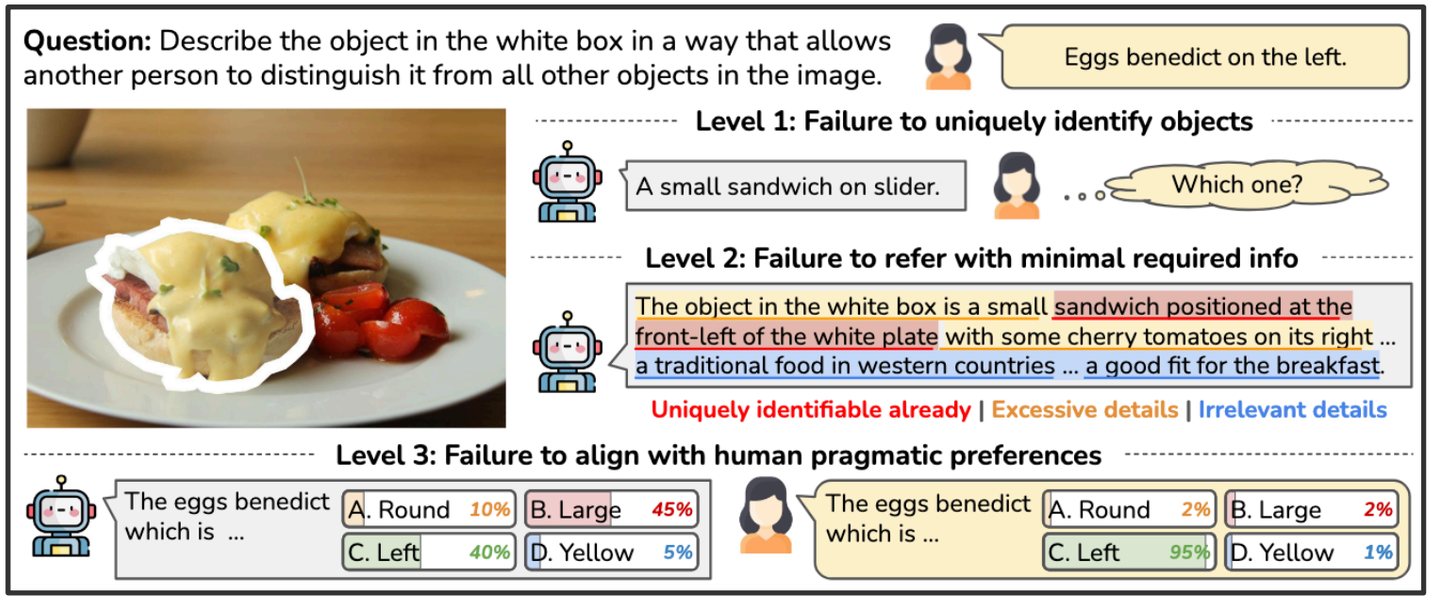
Simple Multi-Agent Scenario: Reference Game

- Very simple objective and information asymmetry, and very easy to evaluate

- Yet, it can elicit sophisticated theory-of-mind reasoning about other agents

- Well-studied in cognitive science, psychology, and linguistics since at least the 1960s, **and still a challenge with current models!**





Krauss and Weinheimer 1964, Ma et al. 2025

Conversation as a Multi-Agent Game

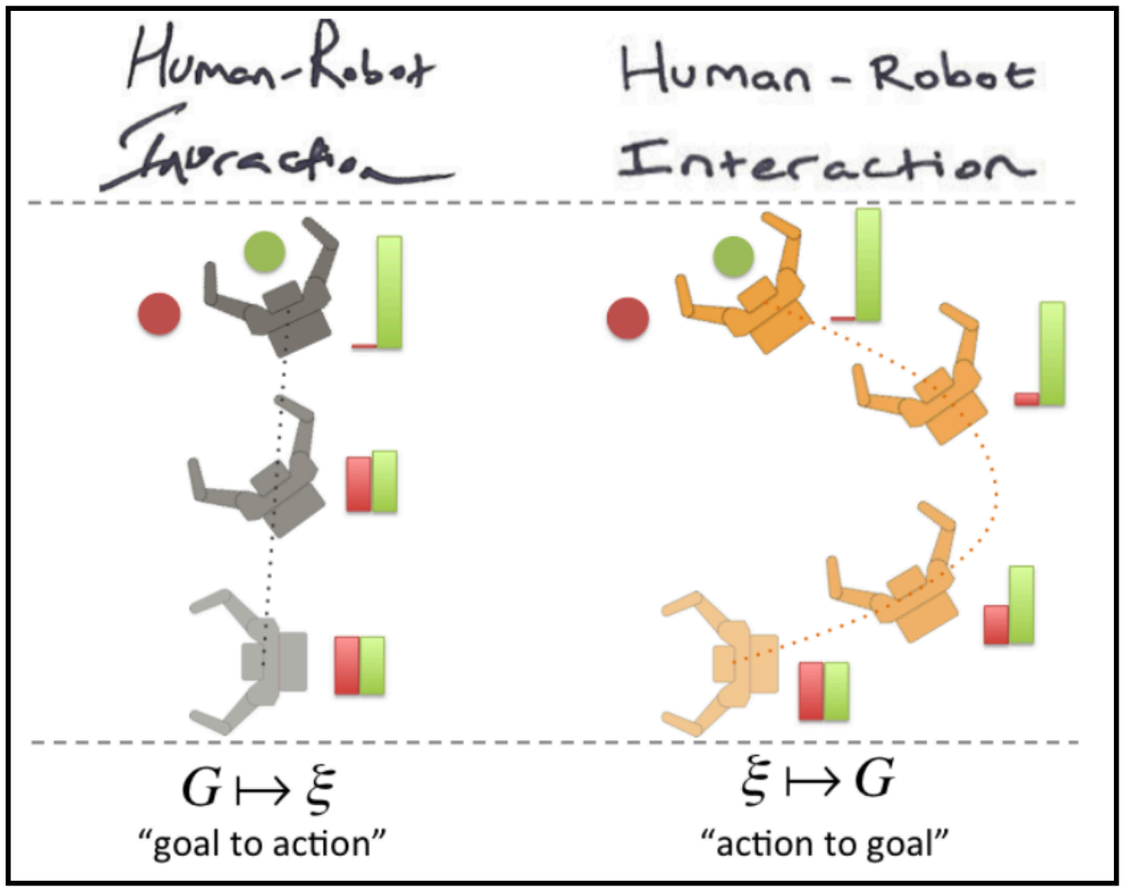
- No immediate physical environment, but we still exist in a social world, with common (or uncommon) knowledge, goals, and values
- One can work towards social goals by using language as action, e.g. through:
  - Education
  - Persuasion
  - Hate speech and dogwhistles

Applications of Interactive Language Systems

Human-computer interaction

- Agents should take actions (including language use) that are maximally informative to users with respect to their beliefs, goals, and actions

- Agents should model and accommodate their users

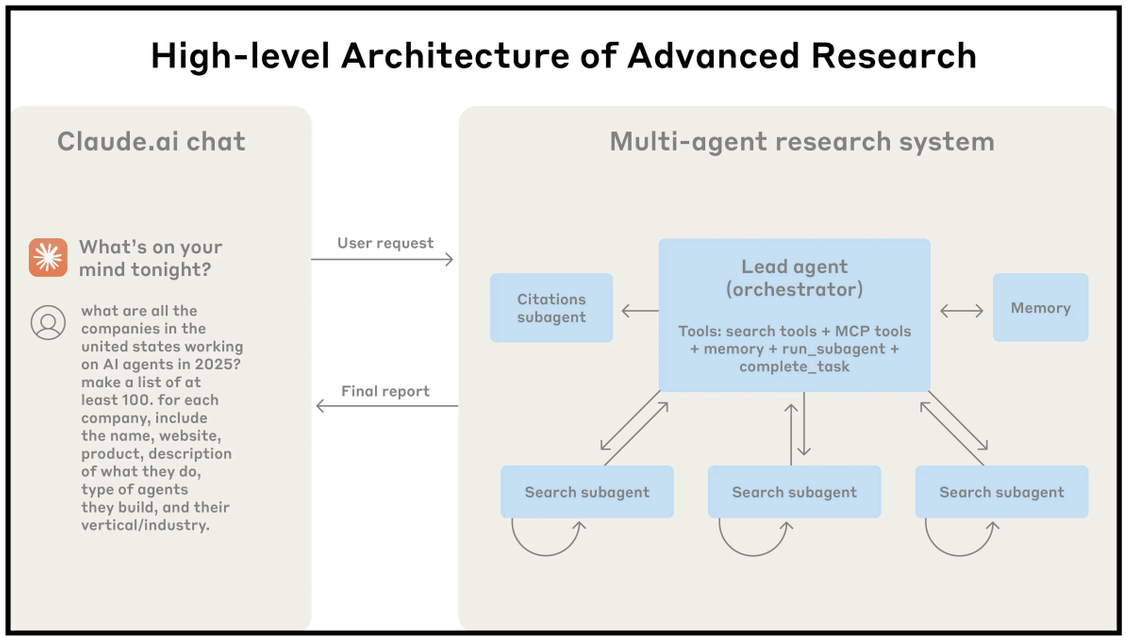


Legibility (Dragan et al. 2013)

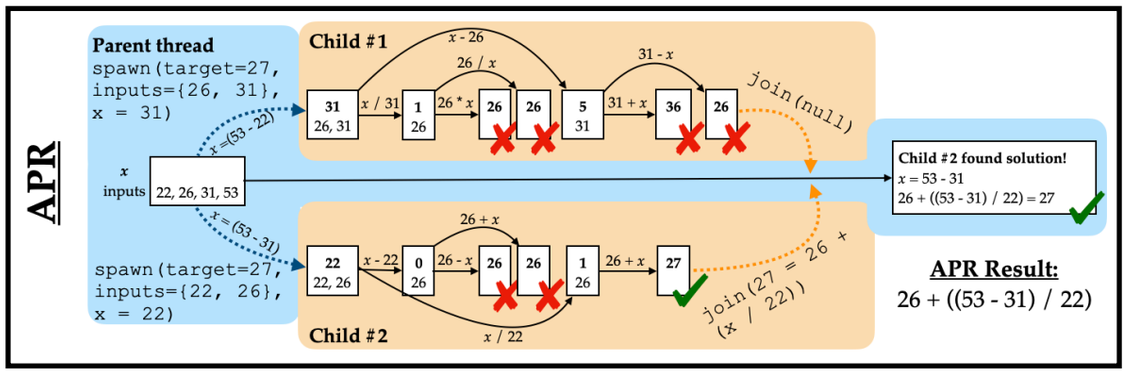
Applications of Interactive Language Systems

Task decomposition and distribution

- Break up difficult tasks into smaller subtasks
- Agents specialize in different subtasks
- Agents can communicate with each other
- Achieve more efficient inference and better utilization of limited context



Multi-agent research (Anthropic 2025)

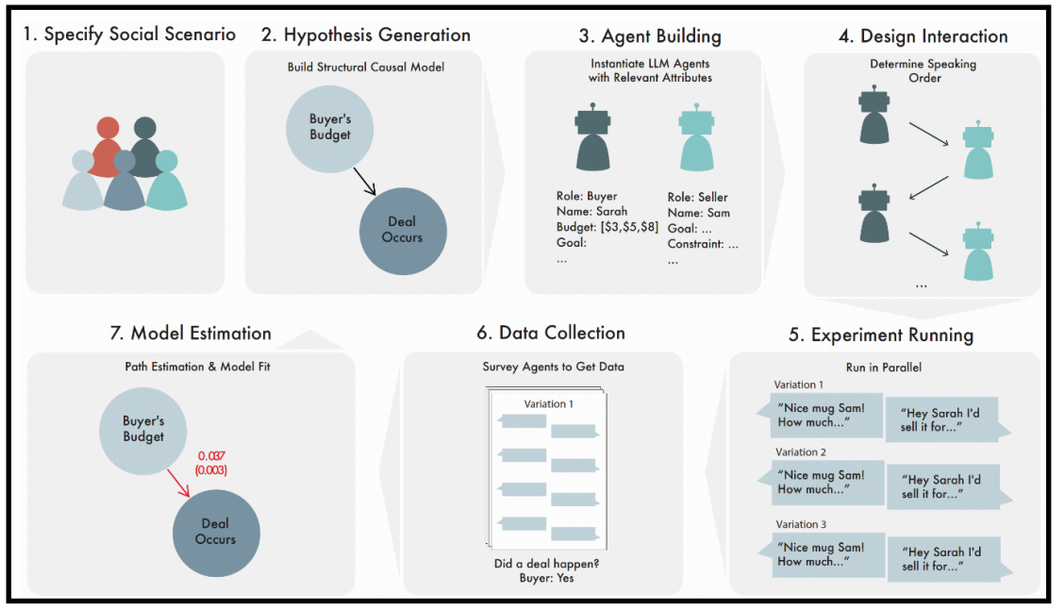


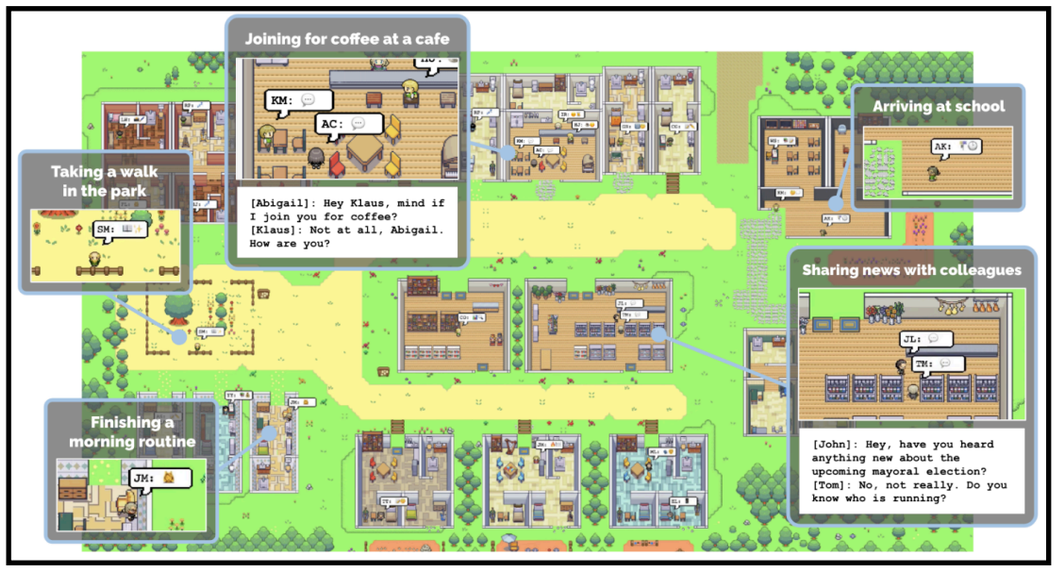
Adaptive parallel reasoning (Pan et al. 2025)

Applications of Interactive Language Systems

Insights into social systems

- Can we replace human subjects research with multi-agent simulations?
  - Can we model individual people/“personas”?
  - Can we model their interactions
  - Can we make conclusions about how, e.g., policies will influence human behavior?
- Big caveat: LLMs flatten social groups and replicate stereotypes



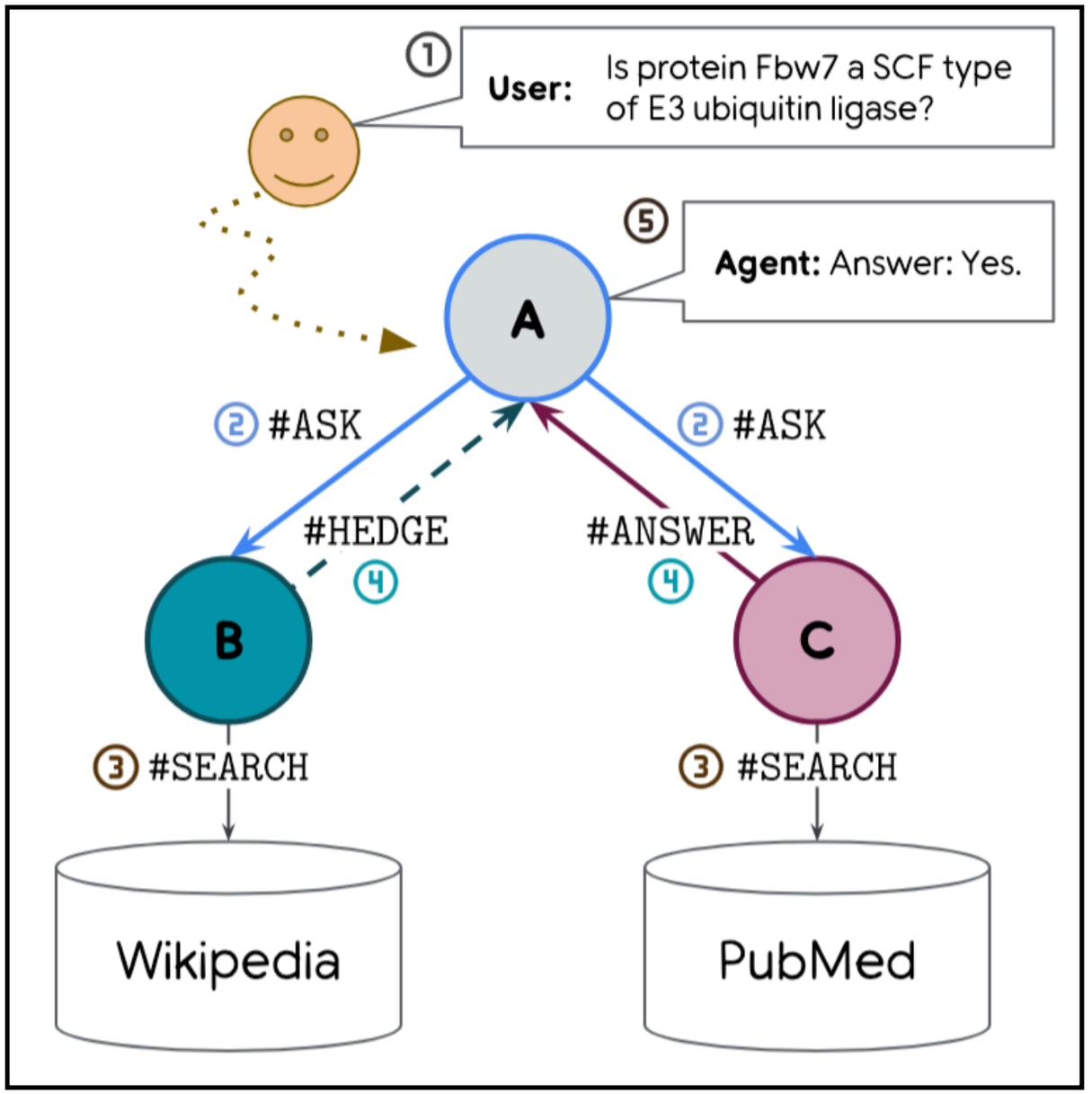


Manning et al. 2024, Park et al. 2022—2025, Meister et al. 2025, Kapania et al. 2025

Applications of Interactive Language Systems

Learning to interact by interacting with oneself

- Can we design games that elicit certain kinds of language use, and learn through self-play?
- Collaborative self-play for linguistic confidence calibration
  - Agent A has no tools, and needs help from B or C, and will choose the answer of the one who sounds most confident
  - C's tool is useful; B's is not
  - Reward the agents if A answers the question correctly — incentives agents to express uncertainty based on information they get from tool use

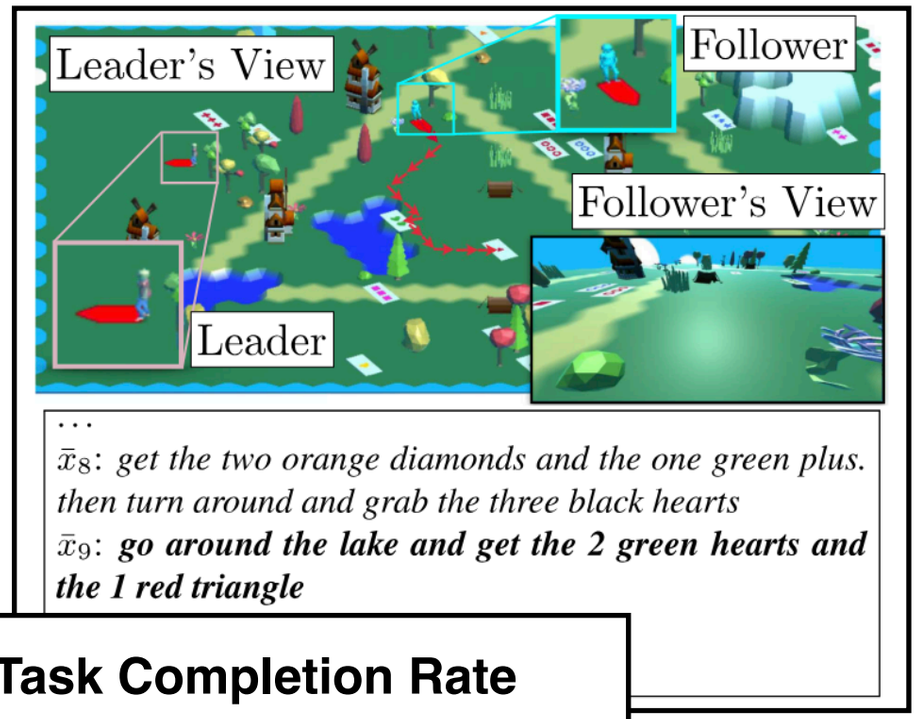


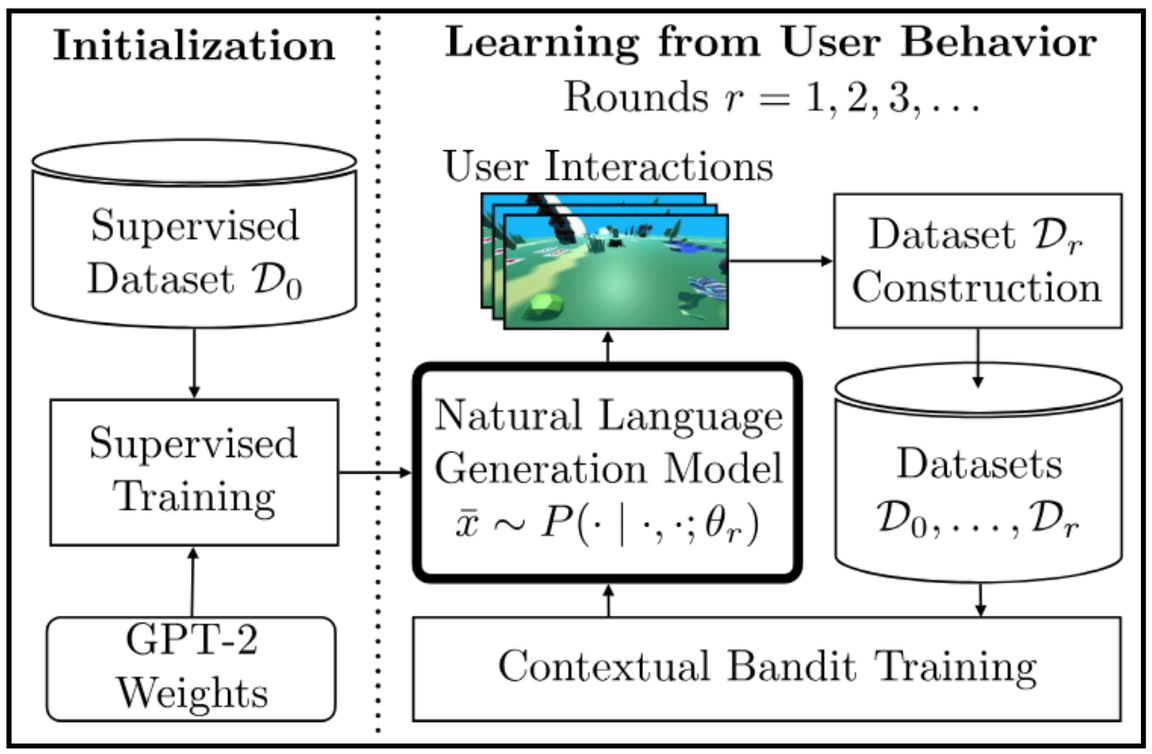
Eisenstein et al. 2025

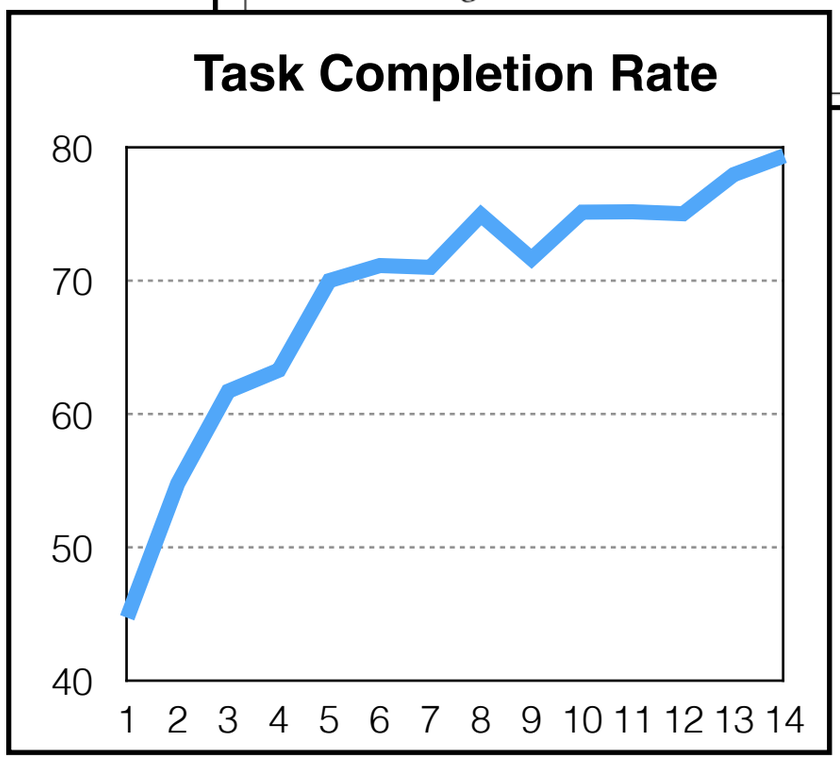
Applications of Interactive Language Systems

Learning to interact by interacting with people

- Deploy learning agent in interaction with people
- Observe user behavior, measure communicative / task success
- Reward agent for achieving success via language-as-action







Kojima et al. 2021, Suhr and Artzi 2023, Gul and Artzi 2024, Tang et al. 2024