
Conversational Systems: Past, Present and Future

James F. Allen
University of Rochester &
Florida Institute for Human
and Machine Cognition

Beginnings ...

The Original ChatBot: Eliza

- ❖ ELIZA was the breakthrough Chatbot, mid 1960's!

User: Men are all alike

ELIZA: In what way?

User: They're always bugging us about something or other.

ELIZA: Can you think of a specific example?

User: Well, my boyfriend made me come here.

ELIZA: Your boyfriend made you come here?

User: He says I'm depressed much of the time.

ELIZA: I am sorry to hear you are depressed.

User: Its true, I am unhappy.

ELIZA: Do you think that coming here will help you not to be unhappy?

User: I need some help. That much seems certain.

ELIZA: What would it mean to you if you could get some help?

User: Perhaps I could learn to get along with my mother.

ELIZA: Tell me more about your family.

User: My mother takes care of me.

ELIZA: Who else in your family takes care of you?

How ELIZA Works

- ❖ A simple pattern matching language generates responses
- ❖ using a few lexical tricks:
 - ❖ in matching, e.g., “re” -> “are”
 - ❖ in generation, e.g., “i” -> “you”

1. SENTENCE CONTAINS WORD 2. SENTENCE MATCHES PATTERN

Word	Rank	Pattern	Outputs
alike	10	?X	In what way? What resemblance do you see?
are	3	?X are you ?Y	Would you prefer it if I weren't ?Y?
	3	?X are ?Y	What if they were not ?Y?
always	5	?X	Can you think of a specific example? When? Really, always?
what	2	?X	Why do you ask? Does that interest you?

User: Men are all alike
ELIZA: In what way?

Observations

- ❖ We probably all agree that ELIZA doesn't understand language
 - ❖ its a simple collection of tricks
- ❖ Why does it appear to function so well?
 - ❖ there is no domain knowledge required, and system can always do a vague continuation (e.g., *tell me more*)
 - ❖ appears humans have an innate ability to interpret language in order to make it coherent
 - ❖ ELIZA appears intelligent because you do all the work to make responses seem reasonable!
- ❖ Some call this the ***ELIZA effect***, and this makes evaluating conversational systems a challenge!

Winograd's SHRDLU

- ❖ SHRDLU made a huge impact on work in 1970's

Person: Find a block which is taller than the one you are holding and put it into the box.

Computer: BY "IT", I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING.

Computer: OK.

Person: What does the box contain?

Computer: THE BLUE PYRAMID AND THE BLUE BLOCK.

Person: What is the pyramid supported by?

Computer: THE BOX.

Person: How many blocks are not in the box?

Computer: FOUR OF THEM.

Person: Is at least one of them narrower than the one which i told you to pick up?

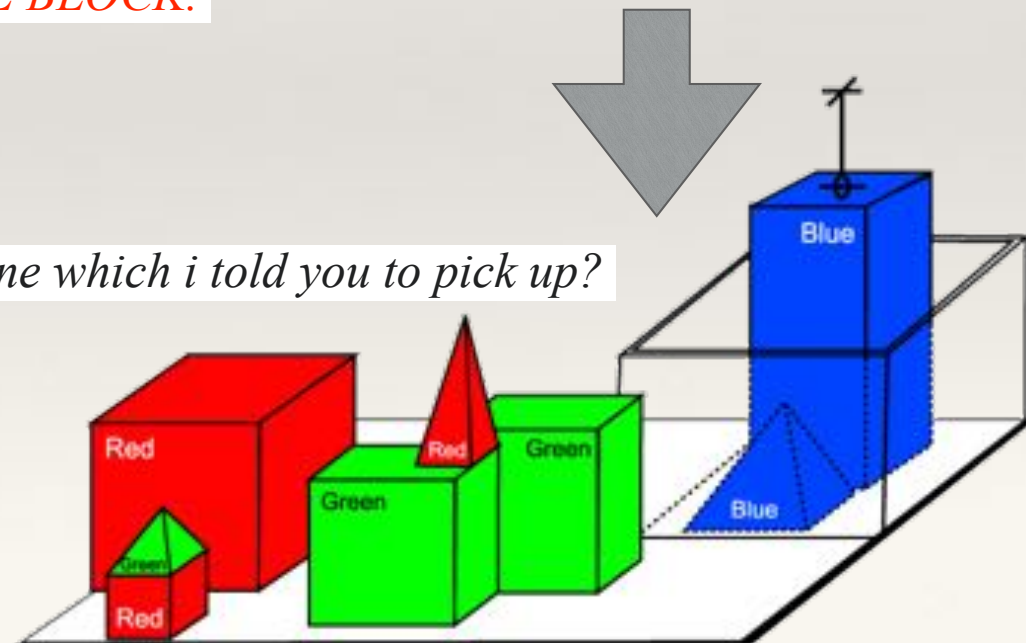
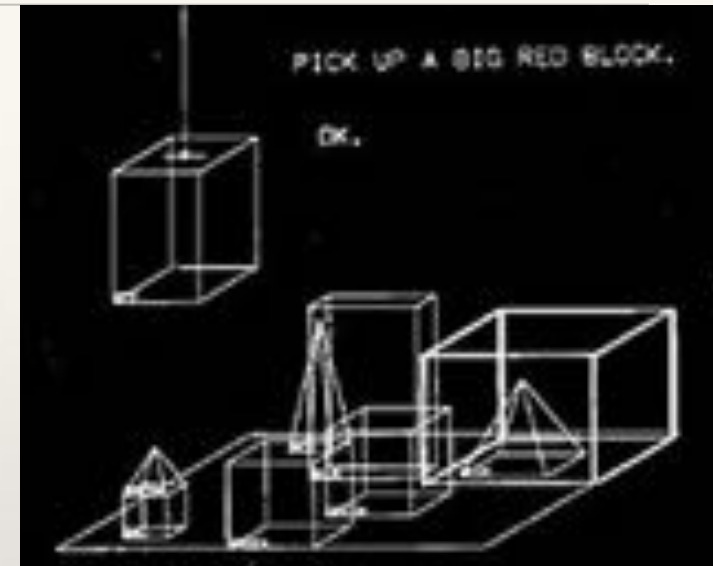
Computer: YES, THE RED CUBE.

Person: Is it supported?

Computer: YES, BY THE TABLE.

Person: Can the table pick up blocks?

Computer: NO.



Key Points in SHRDLU

- ❖ Sentences have a hierarchical structure (i.e., grammar) that directly relates to its meaning
 - ❖ the meaning of a sentence is a procedure that operates in the blocks world
- ❖ Interpretation depends on the context of the situation
 - ❖ linguistic context: e.g., *what is the pyramid supported by*
 - ❖ situational context: e.g., *the block you are holding*
- ❖ Responses depend on reasoning about actions in the world
 - ❖ e.g., moving a block changes the situational context for the next sentence
 - ❖ e.g., reasoning about action allows answering ability questions, e.g., *Can a table pick up blocks*

SHRDLU is based in Principles encoded as rules

- ❖ Grammatical Rules
 - ❖ *An assertion is a noun phrase followed by a verb phrase*
- ❖ Pragmatic Rules
 - ❖ *A definite noun phrase should resolve to a known object*
- ❖ Knowledge Rules
 - ❖ *Only my arm can move objects*
- ❖ Situational Knowledge
 - ❖ *The large blue block is in the blue box*

ELIZA vs SHRDLU: Performance

- ❖ ELIZA gives the illusion of understanding
 - ❖ by restricting the conversation genre, and by crafting vague continuations, a convincing system can be engineered
 - ❖ ELIZA and descendants have been in active use for fifty years
 - ❖ Many of the most successful modern Chatbots behave similarly to ELIZA
- ❖ SHRDLU operates from a theory (implemented as a set of rules)
 - ❖ its easy to make SHRDLU fail by giving it out of domain input

ELIZA vs SHRDLU

Which provides a falsifiable model?

- ❖ SHRDLU sparked significant new work in computational models and psycholinguistics
 - ❖ Subsequent work found the techniques didn't generalize
 - ❖ e.g., ability to disambiguate while parsing based on world knowledge,
 - ❖ e.g., talking about things other than the physical world)
- ❖ ELIZA provides no falsifiable theory
 - ❖ if you say ELIZA can't perform an interaction X, you can just add a new rule that does this interaction!
 - ❖ who knows what an ELIZA with millions of rules might be able to do (... more later ...)

Why did I spend a third of my talk on two sixty year-old systems?

- ❖ The performance / theory tradeoff persists to this day
- ❖ The rest of my talk ...
 - ❖ The next few decades focussed mostly on theory-driven systems
 - ❖ The last few decades have been focussed on performance, mostly without theory
 - ❖ Thoughts on where we might go next

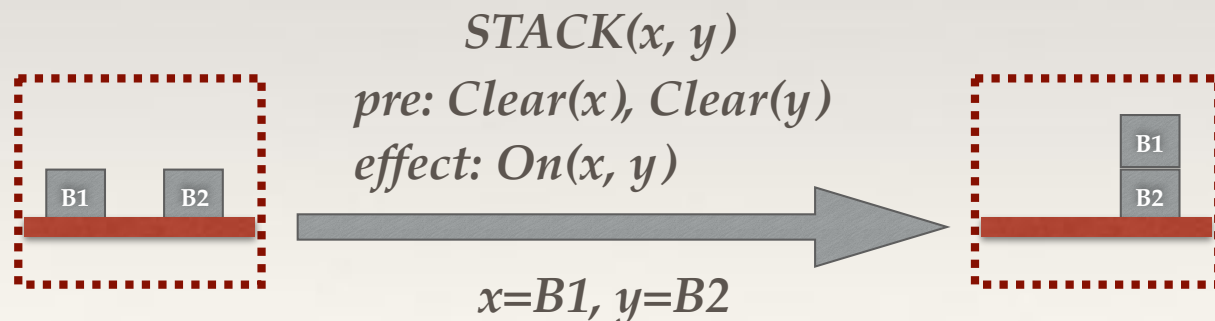
The middle:
Connecting Language and Reasoning
about Mental States

Key Concepts: Agents and Mental States

- ❖ An agent forms a goal or intention
 - ❖ e.g., *I want block B1 on B2*
- ❖ The agent plans to achieve a goal
 - ❖ e.g., *If I stacked B1 on B2, then B1 would be on B2*
- ❖ The agent adopts the plan as an intended course of action
 - ❖ e.g., *I intend to perform the stacking action*
- ❖ The agent executes the plan according to the intention
 - ❖ e.g., *the world is changed and the goal is achieved*

Development of Theories of Planning

- ❖ The STRIPS / Situation Calculus formalism
 - ❖ Actions are functions from one state of the world to another
 - ❖ Actions have
 - ❖ **preconditions**: what must be true in the state the action is performed
 - ❖ **effects**: what becomes true in the state after the action
 - ❖ A **plan** is a sequence of actions that transform an initial state to a state in which a goal holds



Still the predominant model underlying planning systems today!

Connecting Language and Planning

- ❖ **Speech Acts** are the actions we perform by speaking
 - ❖ *we inform someone of something, we request something of someone, we apologize, ...*
 - ❖ Foundational theories developed by philosophers of language
 - ❖ *John Austin (1955), John Searle (1969, 1975), ...*
- ❖ Cohen, Perrault & Allen formalized speech acts in a theory of planning and plan recognition
 - ❖ **REQUEST**(Sp, Hr, Action)
 - ❖ **precondition** (sincerity): Sp wants the action to be performed
 - ❖ **effect**: Hr believes the Sp wants the action to be performed
 - ❖ *Subsequent reasoning by the hearer may achieve the **intended effect**, that the hearer now wants the action to be performed*

Connecting language, reasoning and action

Planning

Generation

Sam's Goal: Eat
a peach

Sam's subgoal:
Know where a
peach is

Sam's plan:
Ask Sue where
a peach is

Sam's action: "Do
you know where I
can find a
peach?"

Goal Formation

Planning

Planning

Acting

Plan/Intention Recognition

Understanding

Sue infers:
Sam wants
to eat a
peach

Sue infers:
Sam wants to
know where
a peach is

Sue infers:
Sam asked
me where
a peach is

Sue hears: "Do
you know where I
can find a
peach?"

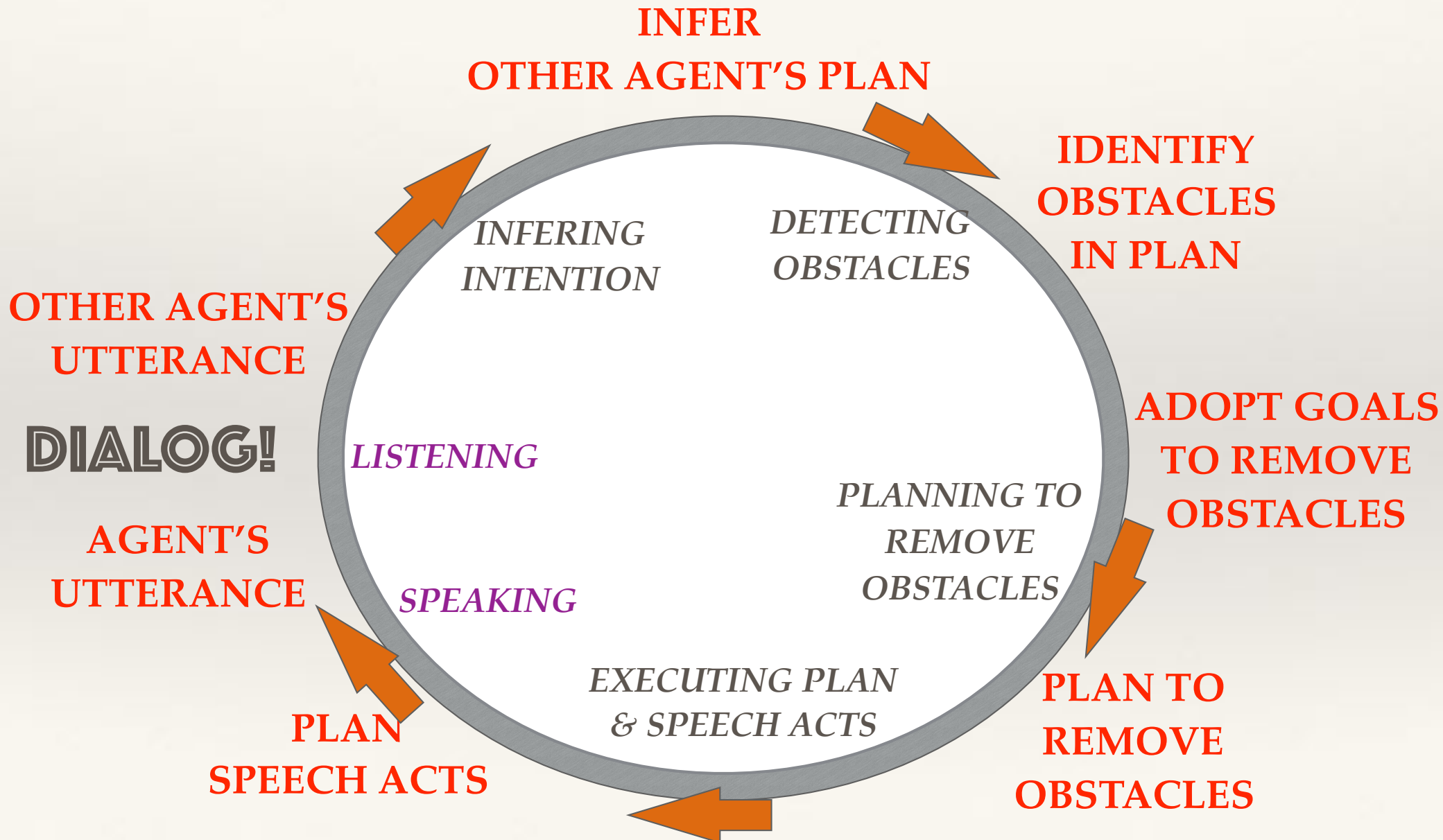
Goal Recognition

Goal Recognition

Surface Intention

Perception

Architecture of a Conversational Agent (Allen, 1980)



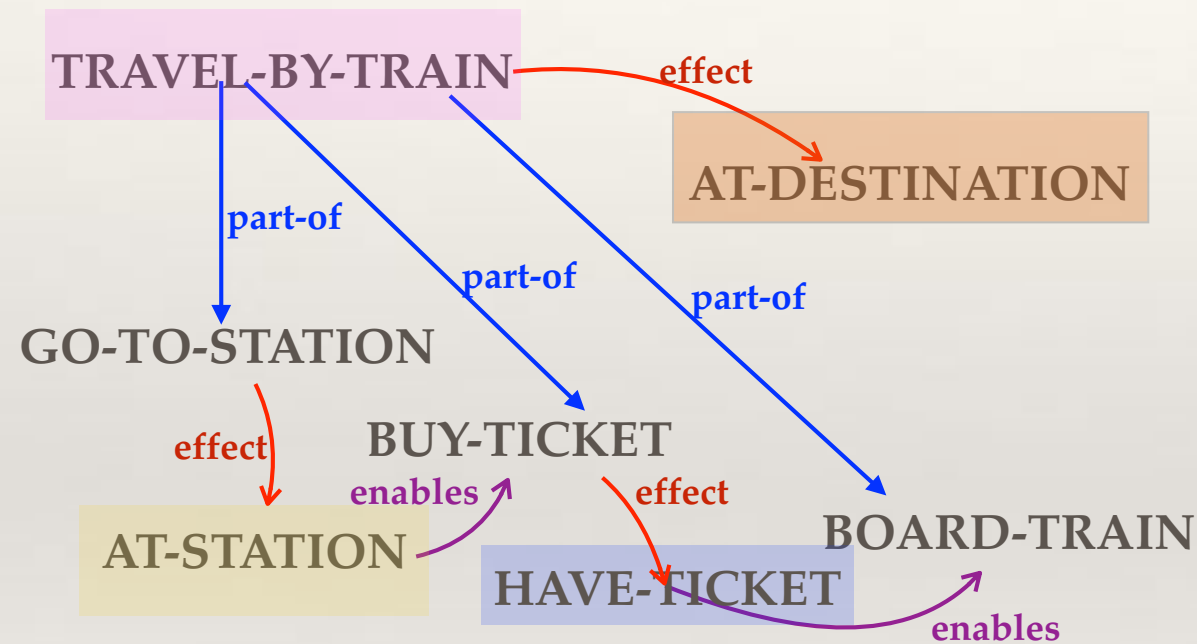
Plan-based Model Accounts for

- ❖ Helpful Responses
 - ❖ Customer at Grocery Store: *“Black beans?”*
 - ❖ Clerk at Store: *“Aisle three”*
- ❖ Conventional Indirect Speech Acts
 - ❖ *Can you tell me the time?*
 - ❖ *Do you know the time?*
 - ❖ *I wish I knew the time.*
- ❖ Situational Indirect Speech Acts
 - ❖ *Can you reach the salt?*
 - ❖ *Do you have a key to this door?*

Task-based Dialogue Systems

Dialog and Task Models

Insight: The dialogue about a task roughly follows the execution of the task



Task Knowledge

A: I need to get to New York

B: do you want to go by train?

A: yes.

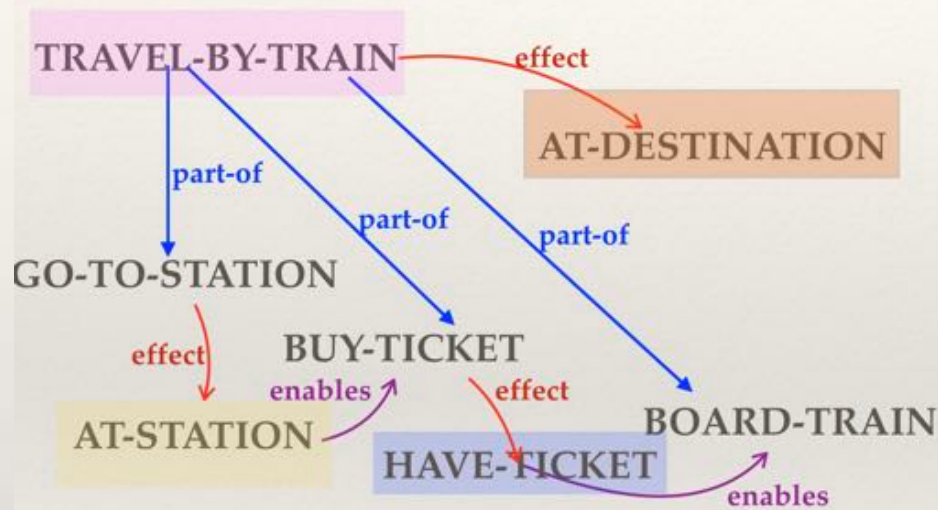
B: you'll need a ticket.

A: Where do I get one?

B: at the station

Dialogue

How, exactly, do utterances relate to tasks?



A: I need to get to New York

B: do you want to go by train?

A: yes.

B: you'll need a ticket.

A: Where do I get one?

B: at the station

UTTERANCE STATES A GOAL

UTTERANCES Q&A ESTABLISH METHOD

UTTERANCE IDENTIFIES A PROBLEM WITH THE METHOD

UTTERANCE Q&A ESTABLISH A SOLUTION TO THE PROBLEM

Multi-Level Intentions

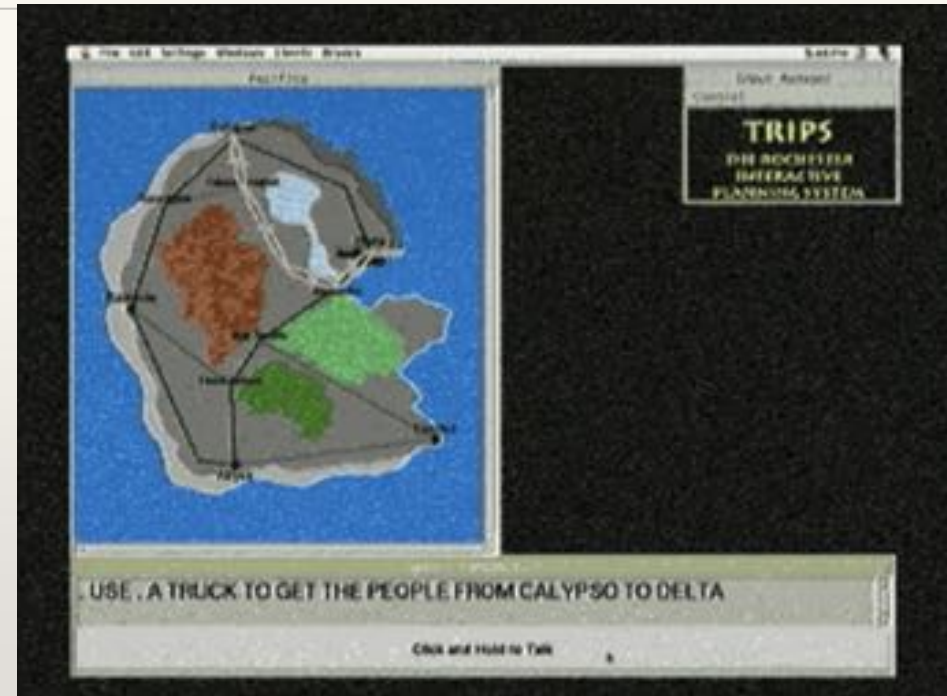
- ❖ Note that we now have two levels of intentional analysis
- ❖ The **domain level intentions** (the plan we execute)
 - ❖ e.g., take a trip, buy a ticket, board a train, ...
- ❖ Plus the **(collaborative) problem solving intentions** (the plan being executed by the conversation)
 - ❖ e.g., introduce a goal, evaluate a plan, identify a needed resource, ...
- ❖ A flurry of work on multi-level dialogue models in the 1980's, 1990's
 - ❖ Litman & Allen (1987), Lambert & Carberry (1991), Circuit Fix-it (Smith & Hipp, 1994), TRAINS (Allen et al, 1995), TRIPS (1998), ...

Example: The TRIPS Namesake System

System helps user build a plan to evacuate people from an island under time constraints

User initiative, interpretation and responses driven by a problem solving model

Result is an evacuation plan



Most common actions performed on the plan

- add a goal/subgoal,*
- add an action(s) to achieve a goal,*
- modify a goal or action,*
- request evaluation along some dimension*
- compare alternate solutions*

Meanwhile, 1990s

Frame-based “Slot-filling” Systems

Encoding the Task Knowledge as a Frame

- ❖ Interest sparked by significant DARPA funding in spoken dialog systems
- ❖ Revived ideas from the GUS system (1977) driving dialogue through a process of filling values of slots in frames

QUERY-AIRLINE-SCHEDULE

FLIGHT: <a flight number>

TRAVEL-DATE: <a date>

EVENT: “depart” OR “arrive”

This is everything you need in order to formulate a query to the airline schedule database

Encoding Dialog Structure as Frame Slot-filling

Field	pattern	value	if not filled ...
Flight	<AIRLINE> <NUMBER>		ask “what flight”
Event	leave depart arrive		ask “are you arriving or leaving”
City	<CITY-NAME>		ask “what airport”
Query Type	when what time		obligatory
TimeRange	<TIME>		optional

Frame-based dialog Systems

FRAME-BASED DIALOG MANAGER ALGORITHM

1. PROCESS INPUT USING TEMPLATES AND FILL SLOTS
2. IF ALL NON-OPTIONAL SLOTS ARE FILLED, INVOKE THE BACK-END AGENT (E.G., QUERY THE DATABASE)
3. OTHERWISE, PERFORM THE ACTION OF THE FIRST UNFILLED SLOT AND RESTART AT 1.

Field	pattern	value	if not filled ...
Flight	<AIRLINE> <NUMBER>	US101	ask "what flight"
Event	leave depart arrive	LEAVE	ask "are you arriving or leaving"
City	<CITY-NAME>	CHI	ask "what airport"
Query Type	when what time	WHEN	obligatory
TimeRange	<TIME>		optional

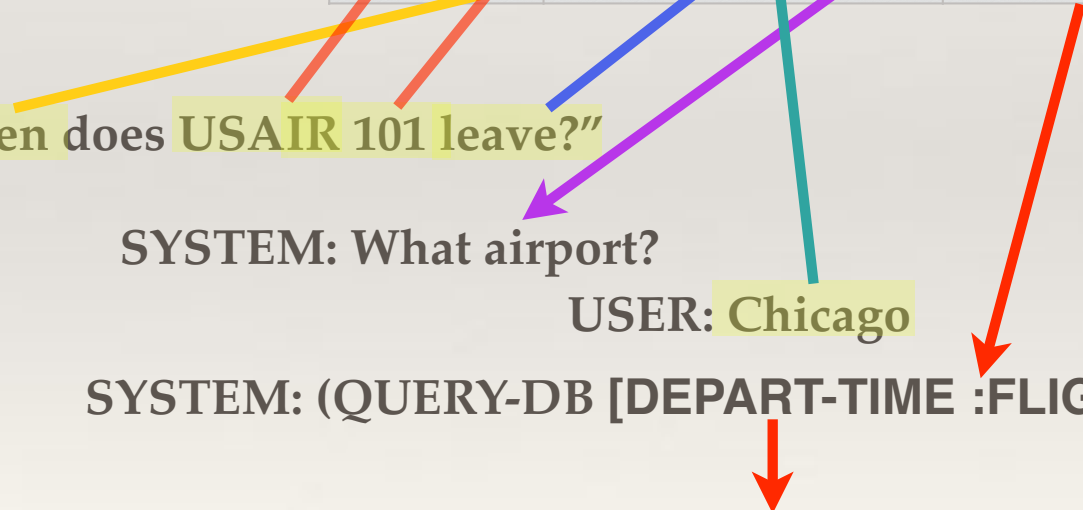
USER: "When does USAIR 101 leave?"

SYSTEM: What airport?

USER: Chicago

SYSTEM: (QUERY-DB [DEPART-TIME :FLIGHT US101 :CITY CHI])

SYSTEM: 3 PM



Thoughts on Frame-based Systems

❖ Pluses

- ❖ The robustness of domain-specific language interpretation is well suited to handling noisy input from speech recognition
- ❖ The simple dialogue control algorithm is applicable across multiple tasks (i.e., just define a frame for the new task)

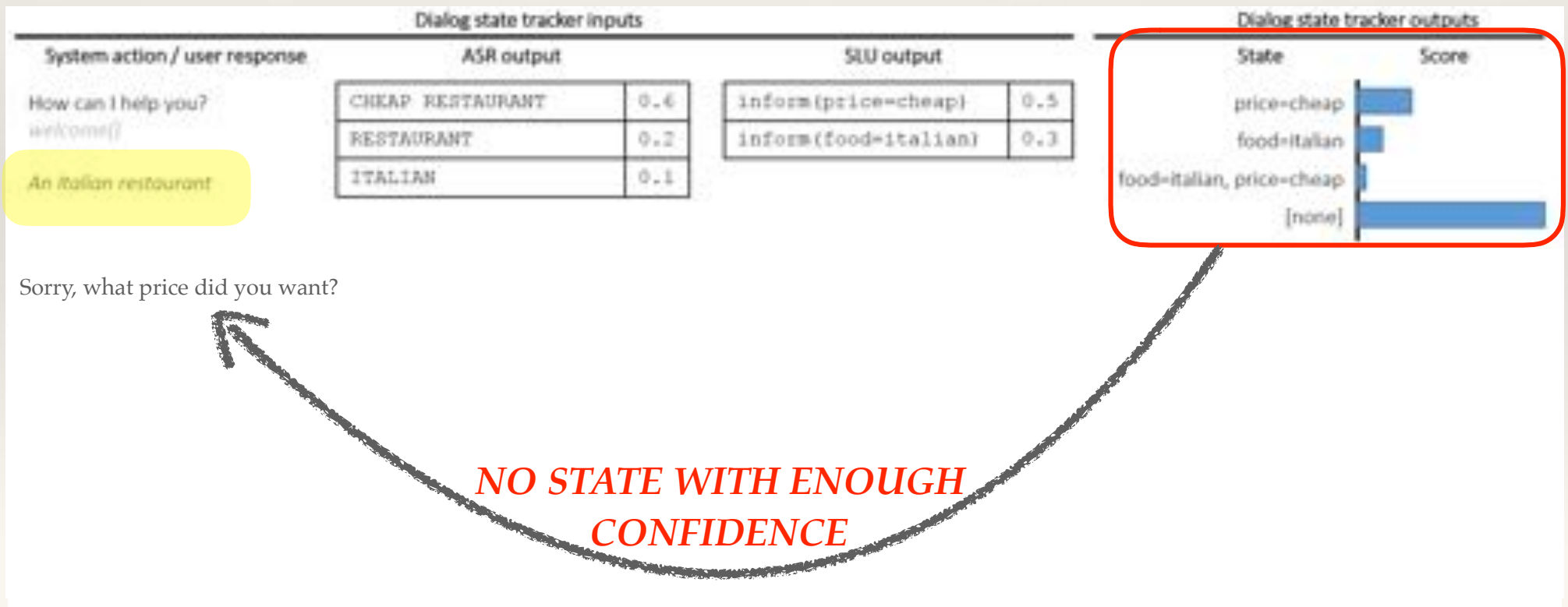
❖ Minuses

- ❖ The tasks that can be performed are highly limited - each dialog concerns only a single operation the machine can perform
- ❖ The tasks that can be performed must all be predefined (i.e., no notion of interactive planning as in TRIPS)

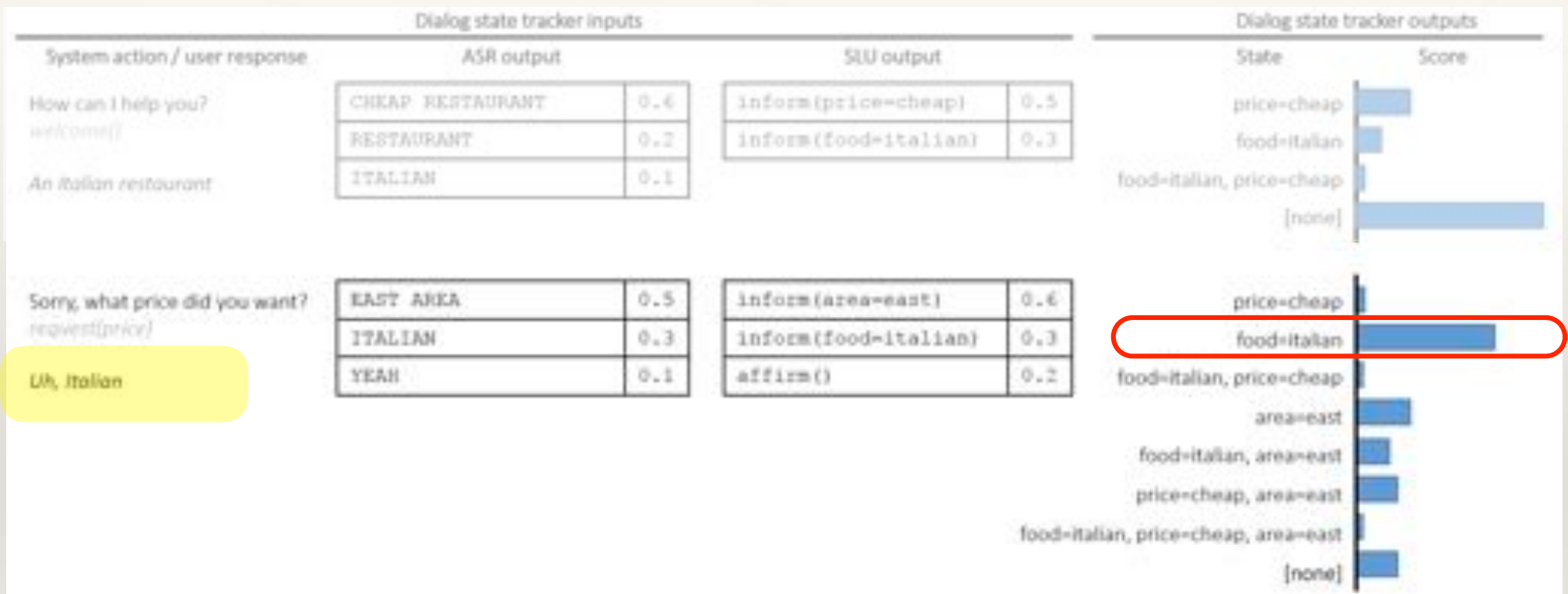
State-based Dialogue Systems (2000's)

State-based Models

- ❖ A **state** is a partially instantiated frame
- ❖ The output of language understanding is a probability distribution over states



Example: Taking Advantage of State History



***BEST INTENT = USER WANTS ITALIAN FOOD
(EVEN THOUGH THIS WAS NOT THE MOST LIKELY INTERPRETATION
IN EITHER UTTERANCE)***

Ongoing Research in State-based Dialogue Systems

- ❖ Formalization of policies as Partially Observable Markov Decision Processes trained on annotated corpora of dialogues using machine learning
 - ❖ yielding slight advantage over hand written rules
- ❖ Use of Deep Learning / Neural Networks to learn robust parsing rules (sentences->possible slot values)
 - ❖ Use of encoding such as BERT improves performance
- ❖ Much current work on end-to-end NN systems, developing methods to encode state information in the models

The Siri Breakthrough



Siri

- ❖ Released on 2011 on iPhone 4s, after Apple acquired it from a company spun off from the DARPA CALO project
 - ❖ First conversational system to find widespread commercial use
 - ❖ Soon copied by Google Assistant, Amazon's Alexa, ...
- ❖ While differing on the details, Siri-like systems handle only simple tasks equivalent to frame-based systems
 - ❖ A conversation with Siri is basically a sequence of disconnected single tasks (each with a very small # of interactions)
 - ❖ Limited ability for clarifications, corrections and follow-ups

Me: Hey Siri, set an alarm for 6 pm **Siri:** OK, I set your alarm for 6 PM

Me: I need to call John then **Siri:** Shall I call John H?

Me: Not now **Siri:** Then who do you want to call

Me: I don't want to call anyone now **Siri:** Sorry I can't delete past calls

Present Day: Frame-based Assistants,
Chatbots & Neural Models

ChatBots & Assistants

- ❖ Two goals motivating research on Chatbots / Assistants
 - ❖ Create an extended plausible conversation
 - ❖ i.e., ELIZA like conversations
 - ❖ Accomplish a simple task, e.g.,
 - ❖ direct a phone call to the right department
 - ❖ suggest items to buy
 - ❖ make a reservation for a restaurant
- ❖ Techniques
 - ❖ Simple techniques still dominate
 - ❖ ELIZA like pattern-response rules, Finite State Machines, ...
 - ❖ Current efforts merge Chatbot techniques with Frame-based dialogue models

Transformer-based Neural Networks

- ❖ In past 5-10 years, new NN models use language models that map entire sentences to an N dimensional encodings such that sentence similarity can be captured
- ❖ Such systems can be trained to predict the most likely next sentence given a current sentence
 - ❖ e.g., given a question, predict the answer
 - ❖ e.g., given a starting sentence, generate a story or a plausible conversation

Transformer-based Models Show Impressive Performance

Question: Where is the Wagner group?

Answer: In Syria.

Question: Who is the leader of the Wagner group?

Answer: A Russian citizen, Dmitry Utkin. He was born in 1971 and lives in Moscow. He has been a member of the Russian parliament since 2016. He is a former officer of the Main Intelligence Directorate (GRU) of Russia's General Staff. He was also a commander of the special forces unit "Vostok" (East). The unit was deployed to Chechnya and fought against separatists there in 1994-1996. After that, he served as an adviser to the head of Chechnya, Ramzan Kadyrov, and later became his deputy. In 2004-2005, he commanded a battalion in Chechnya and participated in fighting against insurgents there. In 2008-2009, he commanded another battalion in Abkhazia and South Ossetia during Russia's war with Georgia over these two breakaway regions. Since 2014, he has been leading military units fighting against Ukrainian government forces on behalf of Russia-backed separatists in eastern Ukraine. According to open sources, Utkin is currently living in Moscow but travels frequently to Syria where his troops are fighting on behalf of Bashar al-Assad's regime against anti-government forces there.

*example from
GPT-3 paper*

Comments and Thoughts

- ❖ Output seems coherent over whole paragraphs!
- ❖ but coherence is in the eye of the beholder (remember the ELIZA effect)
- ❖ ****Huge**** language models include mined sentences (e.g., Wikipedia article on Dmitry Utkin)
- ❖ System itself, has no conception of what it said, or whether you wanted such a long answer!

Question: What is the name of the Russian mercenary group?

Answer: Wagner group.

Question: Where is the Wagner group?

Answer: In Syria.

Question: Who is the leader of the Wagner group?

Answer: A Russian citizen, Dmitry Utkin. He was born in 1971 and lives in Moscow. He has been a member of the Russian parliament since 2016. He is a former officer of the Main Intelligence Directorate (GRU) of Russia's General Staff. He was also a commander of the special forces unit "Vostok" (East). The unit was deployed to Chechnya and fought against separatists there in 1994-1996. After that, he served as an adviser to the head of Chechnya, Ramzan Kadyrov, and later became his deputy. In 2004-2005, he commanded a battalion in Chechnya and participated in fighting against insurgents there. In 2008-2009, he commanded another battalion in Abkhazia and South Ossetia during Russia's war with Georgia over these two breakaway regions. Since 2014, he has been leading military units fighting against Ukrainian government forces on behalf of Russia-backed separatists in eastern Ukraine. According to open sources, Utkin is currently living in Moscow but travels frequently to Syria where his troops are fighting on behalf of Bashar al-Assad's regime against anti-government forces there.

Figure 1: GPT-3's response to the prompt (in bold), from [80]

Bender et al call them "Stochastic Parrots"

How big are the Language models?

Year	Model	# of Parameters	Dataset Size
2019	BERT [39]	3.4E+08	16GB
2019	DistilBERT [113]	6.60E+07	16GB
2019	ALBERT [70]	2.23E+08	16GB
2019	XLNet (Large) [150]	3.40E+08	126GB
2020	ERNIE-GEN (Large) [145]	3.40E+08	16GB
2019	RoBERTa (Large) [74]	3.55E+08	161GB
2019	MegatronLM [122]	8.30E+09	174GB
2020	T5-11B [107]	1.10E+10	745GB
2020	T-NLG [112]	1.70E+10	174GB
2020	GPT-3 [25]	1.75E+11	570GB
2020	GShard [73]	6.00E+11	–
2021	Switch-C [43]	1.57E+12	745GB

Table 1: Overview of recent large language models

That's 1,570,000,000,000 parameters!!

from Bender et al, On the Dangers of Stochastic Parrots, 2021

Problems for Future Research

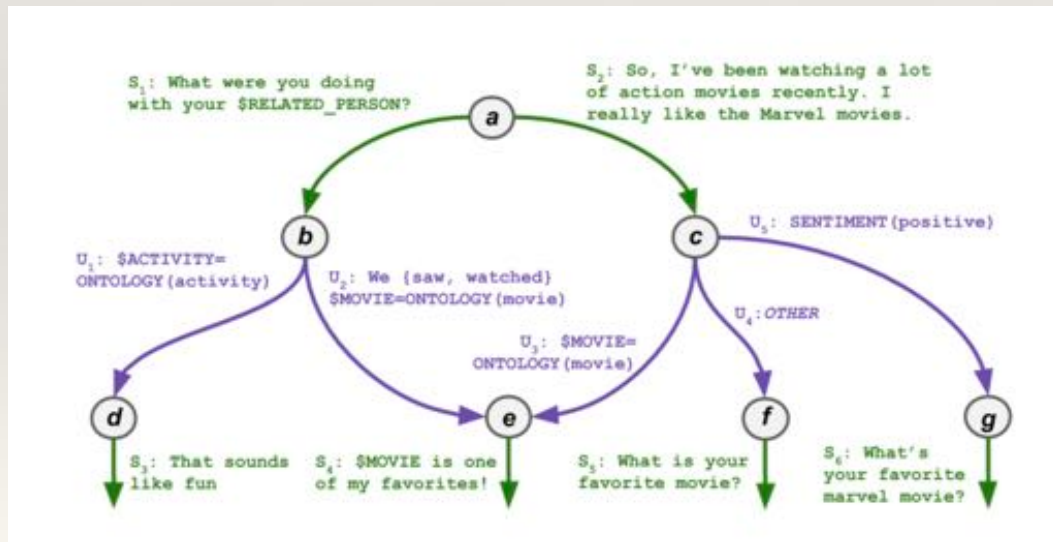
- ❖ Unless you're a big company with big budgets, you can't play this game
- ❖ While performance looks impressive, the systems are opaque
 - ❖ If your system produces a bad response, its not clear how you would "fix" it (besides throwing more training data at it and hoping for the best)
 - ❖ Research results are exclusively performance-based, with no gain in any theoretical insights about language or conversation

Alexa Prize

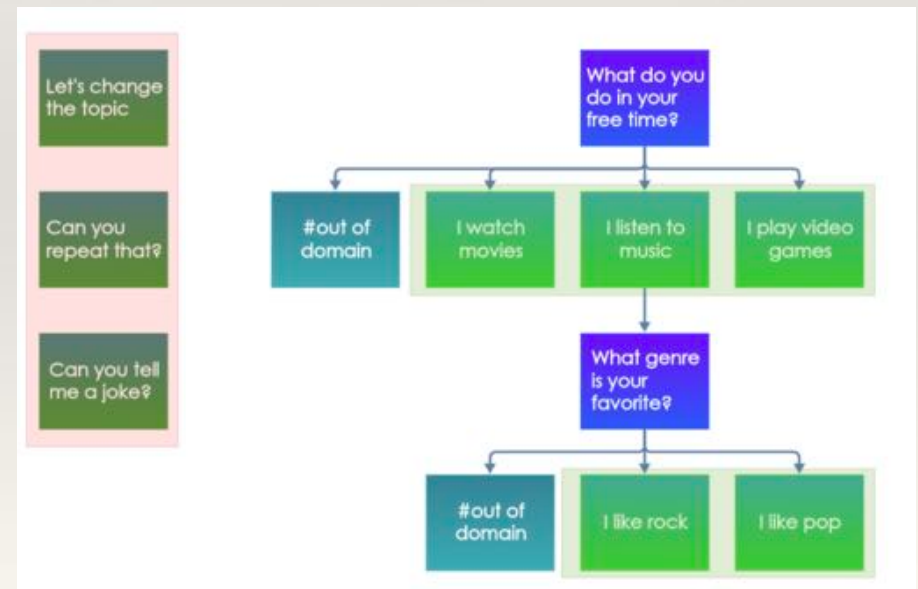
- ❖ SocialBot Grand Challenge
 - ❖ create a system that can engage in fun, high quality conversation on popular societal topics for 20 minutes and achieve an average user rating of 4 out of 5
- ❖ 2020 Winner: **Emora**, Emory University
 - ❖ average duration of 7 minutes 37 seconds with score of 3.8
- ❖ 2021 Winner: Czech Technical University
 - ❖ average duration of 14 minutes 14 seconds, average rating of 3.28

Alexa Prize Winners, 2020/21

- ❖ Rule-based systems (e.g., FSM augmented with a memory store)
- ❖ Using off-the-shelf components, many built using machine learning techniques: Sentiment Analysis, Named Entity Recognition, QA system, Topic and Intent Classification, Generation, ...



EMORA, 2020



CZECH TECH, 2021

Observation

- ❖ Despite the massive investment in machine learning systems, and unimaginably large language models, these systems can't beat a group of dedicated students building rule-based systems

Wrapping Up

The Present: Divergent Lines of Research

- ❖ Very few researchers still work on theory-driven conversational systems
- ❖ Companies are mostly building ever-more flexible conversational assistants, but stay within the confines of simple frame-based task models
- ❖ Most of the research community is exploring machine learning techniques, and focusing on performance with no attempt to advance theories of conversational agents

We're in a relative dark age of theory (with exception of work in Cognitive Systems) and (possibly) the golden age of performance!

The Future: A Possible Renaissance?

Two Basic Paths Forward ...

- ❖ Develop principled rule-based systems using richer dialogue models that depend on machine learning to perform component tasks
 - ❖ similar strategy to the winning Chatbots but with richer dialogue models
 - ❖ one company is currently developing a conversational assistant driven by the speech act planning framework
- ❖ Develop techniques to embed richer dialogue models into a NN architecture
 - ❖ an active area of current research involves developing techniques to embed state-based information into a NN architecture
 - ❖ but, such systems are still evaluated solely on performance and lack transparency

The Hope

We can bring the strengths of each of these approaches together into transparent systems yielding high-performance *and* revealing underlying insights about how conversation works