



# Steps Towards Collaborative Dialogue

**Phil Cohen**

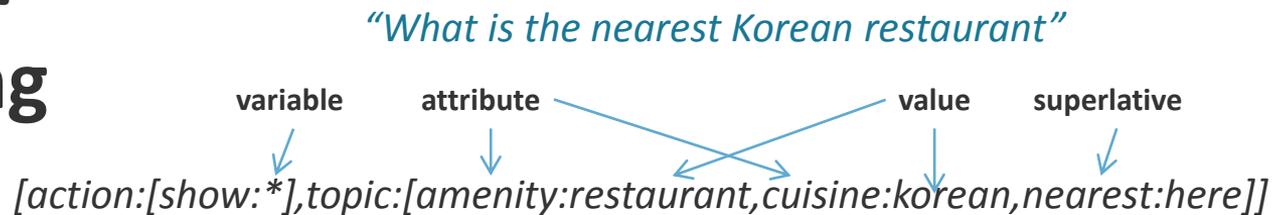
**Chief Scientist, AI &**

**Sr. VP, Advanced Technology**

# Collaborative Interaction

- Part 1: Enabling Technology – Semantic Parsing
- Part 2: How to build a Collaborative Dialogue System

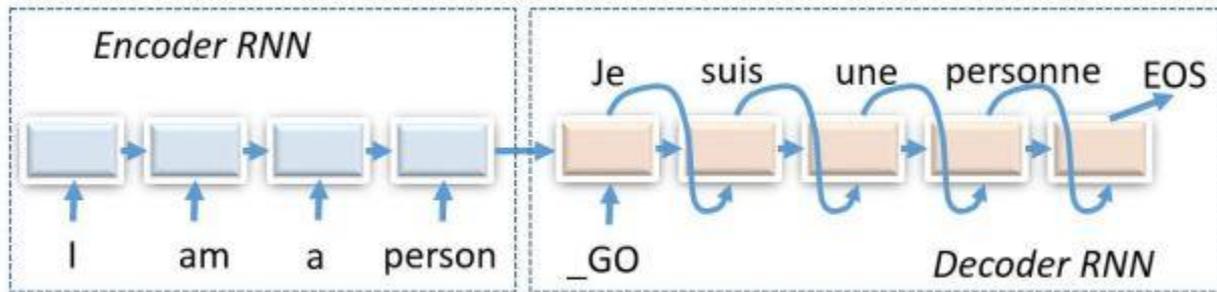
# Progress in NLP: Semantic Parsing



- Process of analyzing a sentence to determine its *semantics*, i.e., its *meaning*
- Meaning is represented as a *logical form (LF)*, a logical language including:
  - » Entities, such as objects in the domain, events, variables, tuples,
  - » Relations such as predicate/argument structures, types, and
  - » Operators, including conjunction, quantification, superlatives, comparatives, aggregation, sequence, conditional, etc.
- Logical forms are *compositional* – meaning of whole is a function of meaning of parts
- Logical forms can be
  - » Vague, with pronouns and referential expressions resolved later via context
  - » Fused with LFs from other modalities
  - » Input to inference
  - » Input to dialogue management subsystems
  - » Mapped to backend data sources,
  - » Executed to retrieve data or invoke APIs,

# Deep Learning Sequence-to-Sequence Models

## Generalizing neural network models for sequence translation



Sequence-to-Sequence (seq2seq) models use a Recurrent Neural Network (RNN) to encode the input sequence, and another RNN to decode it

- » Long Short-Term Memory (LSTM) RNNs can encode/decode long sequences
- » Currently LSTM RNNs achieve state of the art accuracy in *machine translation* and *syntactic parsing*
- » We are using LSTM RNNs to translate English utterances to Logical Forms

# Demonstration of Location Intelligence Application

## Team

### Voicebox Bellevue

- » Phil Cohen, PhD, Chief Scientist
- » Elizabeth Bratt, PhD
- » John Dowding, MSc
- » Arin Goldberg, MSc
- » David McGee, PhD
- » Daniele Terdina

### Voicebox Australia

- » Mark Johnson, PhD, Chief Scientist
- » Hadi Afshar, PhD
- » Long Duong, PhD
- » Dominique Estival, PhD
- » Glen Pink, PhD

Duong et al., Multilingual Semantic Parsing and Code-switching,  
ACL Conf. on Natural Language Learning, 2017

# Opportunity

- Systems can now learn to robustly produce complex logical forms from natural language
- Key is choice of logical forms (LFs) and corpus of utterances paired with LFs
- What should those logical forms be?

# Collaborative Dialogue

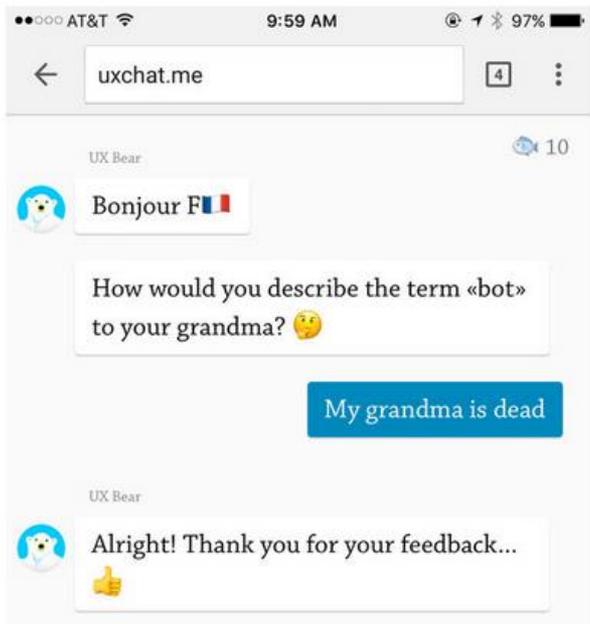
Approaches to dialogue management

What does it take to collaborate?

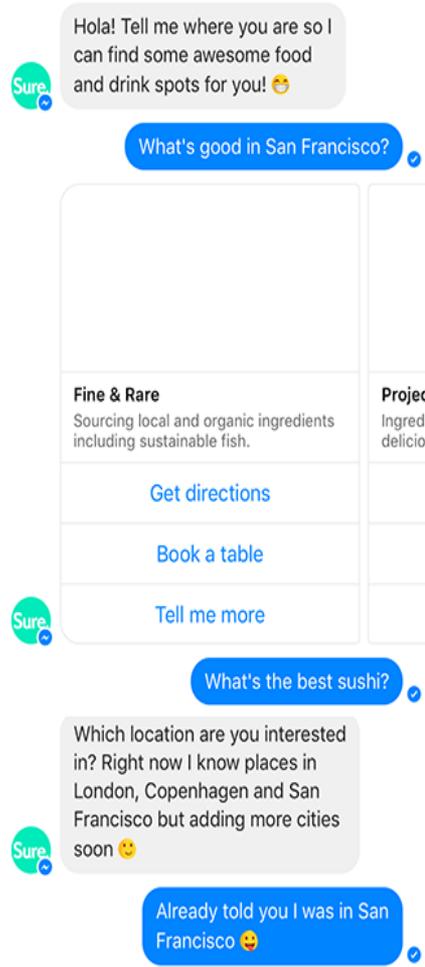
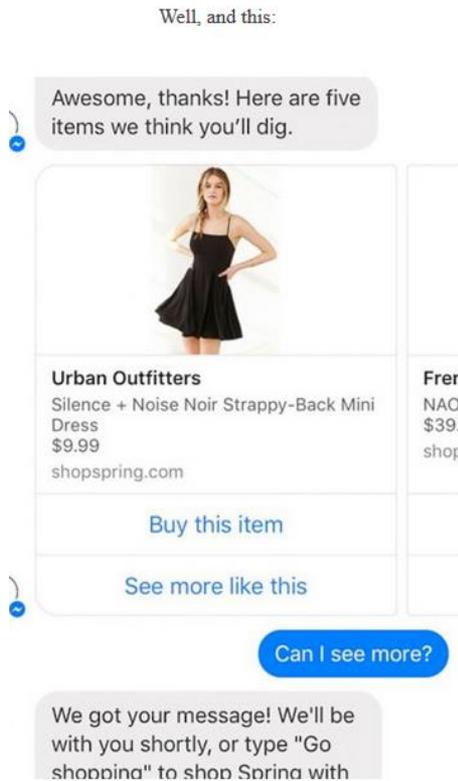
# Current state of dialogue systems

- Chatbots
  - » Eliza (1966) – (poor) imitation of a Rogerian psychotherapist
  - » End-to-end learning stimulus-response from existing dialogue corpora (Parrot-bots)
  - » Dialogue *imitation vs participation*; Bertrand Russell
  - » 30,000 bots on Facebook Messenger; few are used
  - » Woebot (Ng et al.) – Mental health. Vulnerable population. Ethical considerations.
  - » Current research – Learn to interact with knowledge bases
- Finite state dialogue models/controllers –
  - » Sequences of speech acts; used for “authoring dialogue flows”
  - » Foundation for most industry dialogue agents
  - » Large, complex, often rigid. Typically, system initiated dialogue
  - » Best incarnation – Harel State Charts (e.g., Andre et al.)
  - » Typically, hand-built. Compile into them? (compile what?)

# Examples of failures (from: [uxdesign.cc](http://uxdesign.cc).)



Our very own [uxchat.me](http://uxchat.me) pretending it is really listening to what users are saying



# What's missing from this approach

- Human email:
  - » “Take note of ...”
- Google suggested reply:
  - » “thanks for sharing” “good idea”, ...
  - » No use of linguistic coherence items (e.g., use of pronouns)
- Human supplies the *intention*

# Frame-based (aka “slot-filling”) Dialogue Systems

GUS – Bobrow et al., 1977; (also DARPA ATIS systems)

**GUS:** Hello. My name is GUS. I can help you plan a simple trip by air.  
Where do you want to go?

(1) **Client:** I want to go to San Diego on May 28

**GUS:** What time do you want to leave?

(2) **Client:** I must be in San Diego before 10 am

**GUS:** Would you like P.S.A. flight 102 that arrives at 9.15 am?

(3) **Client:** What is the next flight?

**GUS:** Air California flight 310 that leaves at 8.30 am and arrives at 10.00 am

(4) **Client:** I'll take the first one

**GUS:** In what name should I make a reservation, please?

(5) **Client:** Dan Bobrow

**GUS:** I have confirmed the following flight:

P.S.A. flight 102 on Wednesday May 28 from San Jose to San Diego  
that leaves at 7.30 am and arrives at 9.15 am.

What date do you want to return on?

**Flight “frame”**

**Required arguments:**

- Departure-date
- Departure-time
- Destination
- Airline
- Arrival-Date
- Arrival-Time

**Nowadays:**

- NLP for dialogue interfaces is (mostly) “intent classification” + “slot-filling”
- Learn slot-filling dialogue systems from data derived from user simulators (eg., Hakkani-Tur et al)
- Note: Subset of “task-oriented dialogue systems”

# Semantics of utterances about “Slots”

- Utterances about “slots” encode more than content, they also encode **Attitude**
- “the time you want to leave”  $\neq$  the time you can/will leave!
- What about:
  - » *“I want to arrive by 8am, or I will have to go the night before”*
  - » *“If I take the 8am departure, how long will I have to wait in Reykjavik before the flight to Rome*
  - » *“I need to get to Boston on Tuesday. I prefer not connecting in Chicago and hate flying United.”*
  - » *“When you get to the arrivals hall, ask someone where the stand is for the bus to Suwon”*
- These express constraints on slot values
- Can’t just fill a slot.
- What is the language of these constraints (natural and logical)?
- *The only way to express the meanings of these constraints is via a LOGIC*
- *What logic should we use for logical forms? (stay tuned)*

# Back to the Future

## Principles for next generation dialogue systems

- Systems need to understand what you meant (cf Grice, 1957),  
i.e., your intentions
  - » Not just what utterance means
- Systems need to infer why users said what they did
- Systems need to understand users mental states
- Useful systems need to be able to collaborate\*

\*There are certainly nefarious users with whom the system should NOT collaborate

# What is Collaboration?

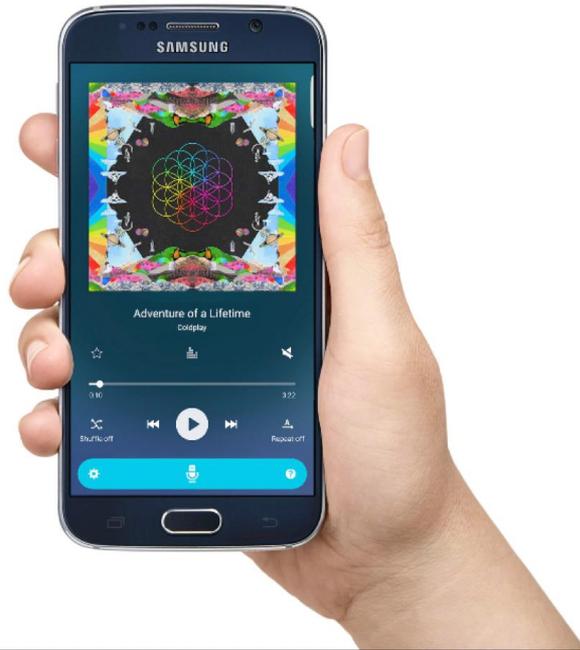
Warneken and Tomasello, 2006



Jointly committed to  
each other's success  
(Cohen and Levesque, 1991;  
Levesque et al., 1990;  
Grosz and Sidner, 1990)

4:17

# Collaborative dialogue driven by plan recognition



- System infers purpose of question is for user to see next concert in region
- System offers to buy tickets
- System informs user that plan will fail (venue is sold out)
- System finds another method to see concert
- System offers to buy tickets to alternative concert for user
- System provides seating chart to get information needed to purchase ticket

# Another dialogue

## actual

- Student:
  - » “Where are the chuck steaks you advertised at 88 cents per pound?”
  
- Butcher:
  - » “how many do you want?”

# Another one

Imagine a Traveler landing in a different country (say, S. Korea) for the first time who needs to connect by bus to her destination. The Traveler has approached an information provider (person or system) at 10:45 pm:

- *Traveler: Do you know when is the next bus to Suwon?*
- *Provider: Sorry, the last bus has left for the evening. You will have to take a bus to Seongnam and transfer to the bus to Suwon. The bus leaves here from bay number 6 at 11:00 pm.*

VS literal answers:

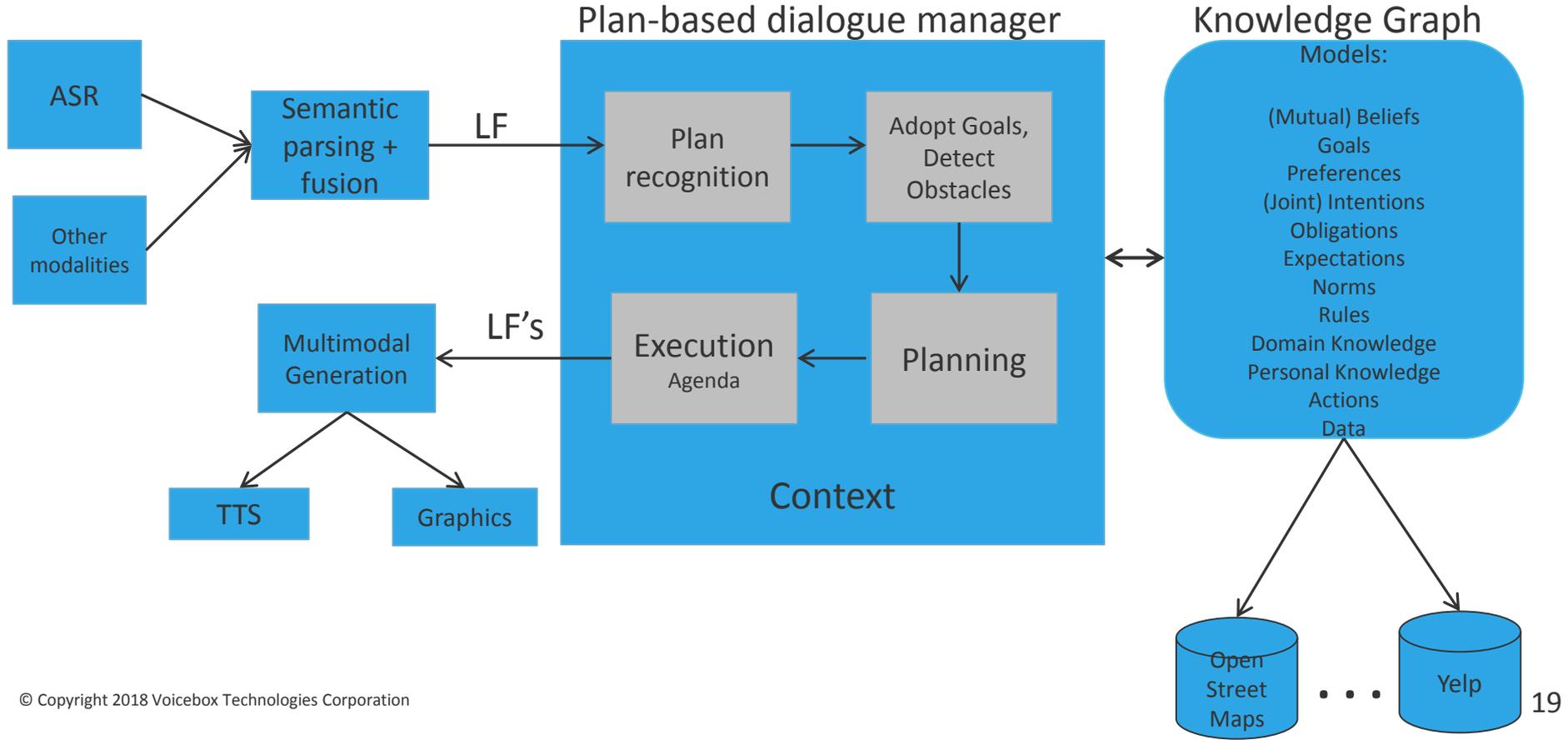
- *Provider: Yes, or*
- *Provider: The next bus leaves tomorrow at 5am.*

# A last collaborative dialogue

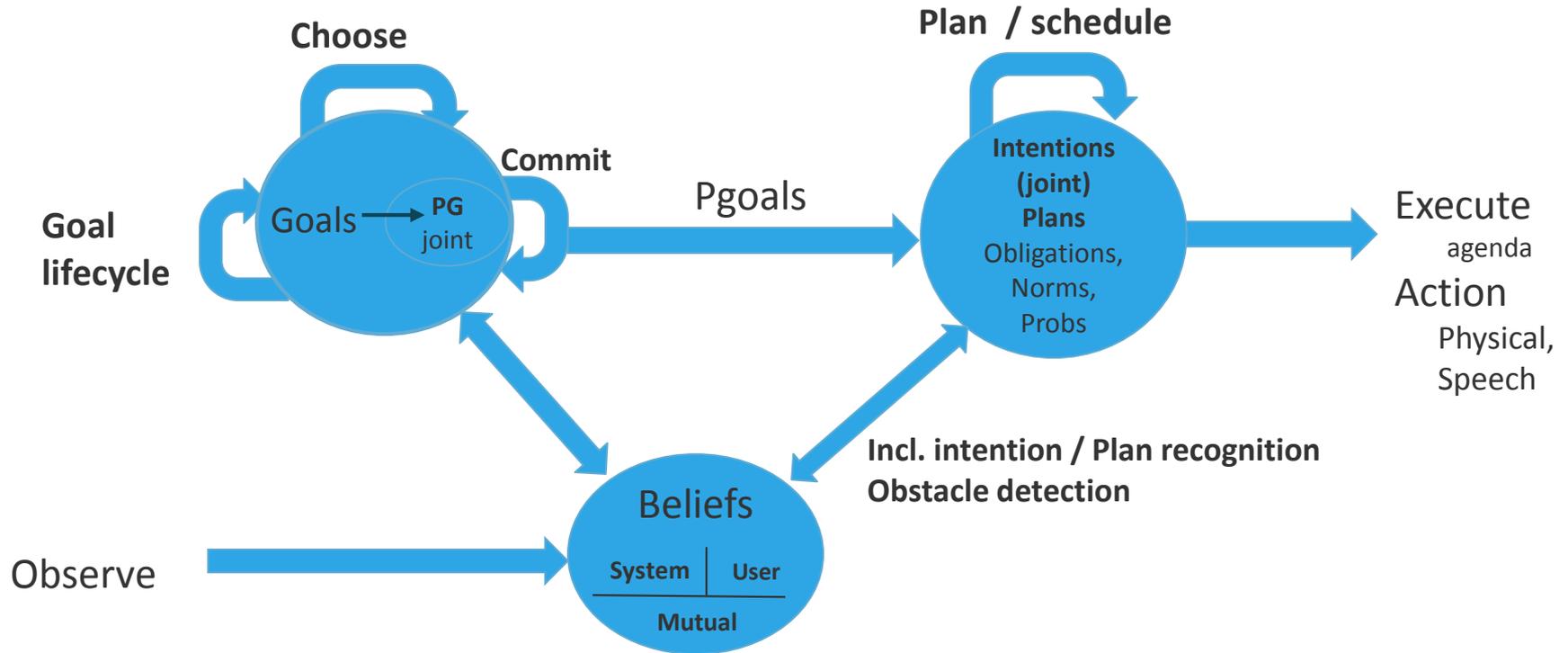
- System provides proactive assistance based inferring on the **User's Plan**
  - » User: *Where is Rogue1 playing tonight?*
  - » System: *It's playing at the Varsity theater at 7:30pm and at the Roxy theater at 9 pm. Would you like me to get you tickets to one of those showings?*
  - » User: *yes, and I'd like to eat at Guillaume before*
  - » System: *The only reservation at Guillaume is at 6:30 pm. There would not be enough time to eat and attend the 7:30 showing. Shall I purchase tickets for the 9pm showing at the Roxy?*
  - » User: *ok*
- System is proactive (offers to buy tickets) and reasons about actions/plans (purchasing tickets, reserving dinner) and schedules (eating; driving; watching)

# Collaborative dialogue

Dialogue management based on plan recognition



# Collaborative BDI Architecture



# Plan-Based Dialogue Modeling

Allen, Bruce, Cohen, Perrault, Levesque, Kumar, Ferguson, Grosz, Pollack, Sidner, Rich, Litman, McRoy, Carberry, Traum, Bretier, Sadek, Bohus,... See also recent literature on epistemic planning (e.g., Muise et al., 2015)

- Same planning/plan recognition algorithms –*physical* and *speech acts*
- Planning of physical acts → reasoning about *physical* states
- Planning of communicative acts → reasoning about *mental* states
- Collaborative dialogue “falls out for free” from a **collaborative belief-desire-intention (CBDI) architecture** that reasons about mental states and speech acts.
- “Frame-based dialogue” “falls out for free” from the CBDI arch. when reasoning about what User knows/needs to know.

# Reasoning about what User is likely to know

- Person 1 hands person 2 this:
- Person 2 asks: “what is this?”
- Person 1: “It’s a USB stick”
- Person 2: “Are you trying to be passive aggressive or something? Of course I know it’s a USB stick!” [he wanted to know what file was ON the stick]
- People are expected to respond based on what others know/believe/intend



# Mental States

## Modal Operators (See Allen&Perrault, 1980, Cohen&Levesque, 1990)

### Multi-modal logic

- **Belief**

- » **bel**(X,P) -- X Believes P

- » **knowref**(X, Var#Type, P) -- Exists Var#Type, **bel**(X, P), st. Var is free in P

- Knowref(john,Num#phone\_number & cell\_phone(bill,Num) )

- John knows Bill's cell phone number vs. John believes Bill has a cell phone #

- "quantifying-in"

- » **knowif**(X,P) -- **bel**(X,P)  $\vee$  **bel**(X,  $\sim$ P)

- John knows whether or not the store is open, vs. John knows the store is (open or not open)

- **PGoal** (Persistent Goal) – **pgoal**(X,P,Q) - can drop if **bel**(X,P)  $\vee$  **bel**(X, $\square\sim$ P), or **bel**(X, $\sim$ Q)

- **Intend**  $\equiv$  **pgoal**(X, do(Agt,Act), Q) – leads to action

- Joint Persistent Goal/Joint Intention/Mutual Belief (for next time) (C&L'91)

# Common Patterns of Plan Inference

Intent / Plan recognition

- Request(Action) → Action<sub>1</sub> → Effect<sub>2</sub> → Action<sub>2</sub>... → Effect<sub>n</sub> (top goal)
- Req(Time of Next concert) → (Date/Loc) → Go to Loc → Attend Concert → Hear musical group
- ↓ Precondition failure
- Check preconditions
  - Have ticket
- On failure, find another plan to achieve goal
  - » Find substitute objects for goals; recommend; edit plan, confirm
  - » Other operations possible -- system plans to overcome obstacles; system plans to achieve goals
- Common precondition failures --
  - » business must be open to conduct commerce in person → find one that is open
  - » Item is in stock → find another business selling item, (e.g., concert venue)
  - » Cannot arrive on time for an appointment → make another appointment; change vendor

# Planning/Plan recognition

- STRIPS-like action representation: Action parameters, Preconditions, Effects, Bodies (hierarchical decomposition), Constraints, Applicability conditions
- **Planning** Rules (backward chaining): **Effect  $\rightarrow$  Action, Action  $\rightarrow$  Precond, (Action  $\rightarrow$  Body)**
  - » Know-value, Reverse a negative state
- **Plan Recog Rules** (forward chaining): **(Precond  $\rightarrow$  Action), Action  $\rightarrow$  Effect, Body  $\rightarrow$  Action**
  - » Normal Activity,
- Subject to: Constraints on parameters, and applicability conds (that cannot be made true)

Potential other methods: plan parsing, Bayes Nets, Action2Vec, abduction... 25

# Example Action

## Actions, parameters, roles, values, value types

The action of **servicing** an auto:

Repairshop services Usr's Auto on Date at Time

Constraint: Repairshop is located at Loc

Precond: Usr has an appointment at Repairshop for Date/Time

Effect: Auto is fixed at Time

# Speech Acts (aka dialogue acts)

Represent speech acts just like physical acts

Inform, informref, informif, offer, request, suggest, confirm, whq, ynq, ...

Speaker **informs** Listener of the referent of **Var#Type**, s.t. **Pred**  
*the Num#phone\_number s.t cell\_phone(john,Num)*

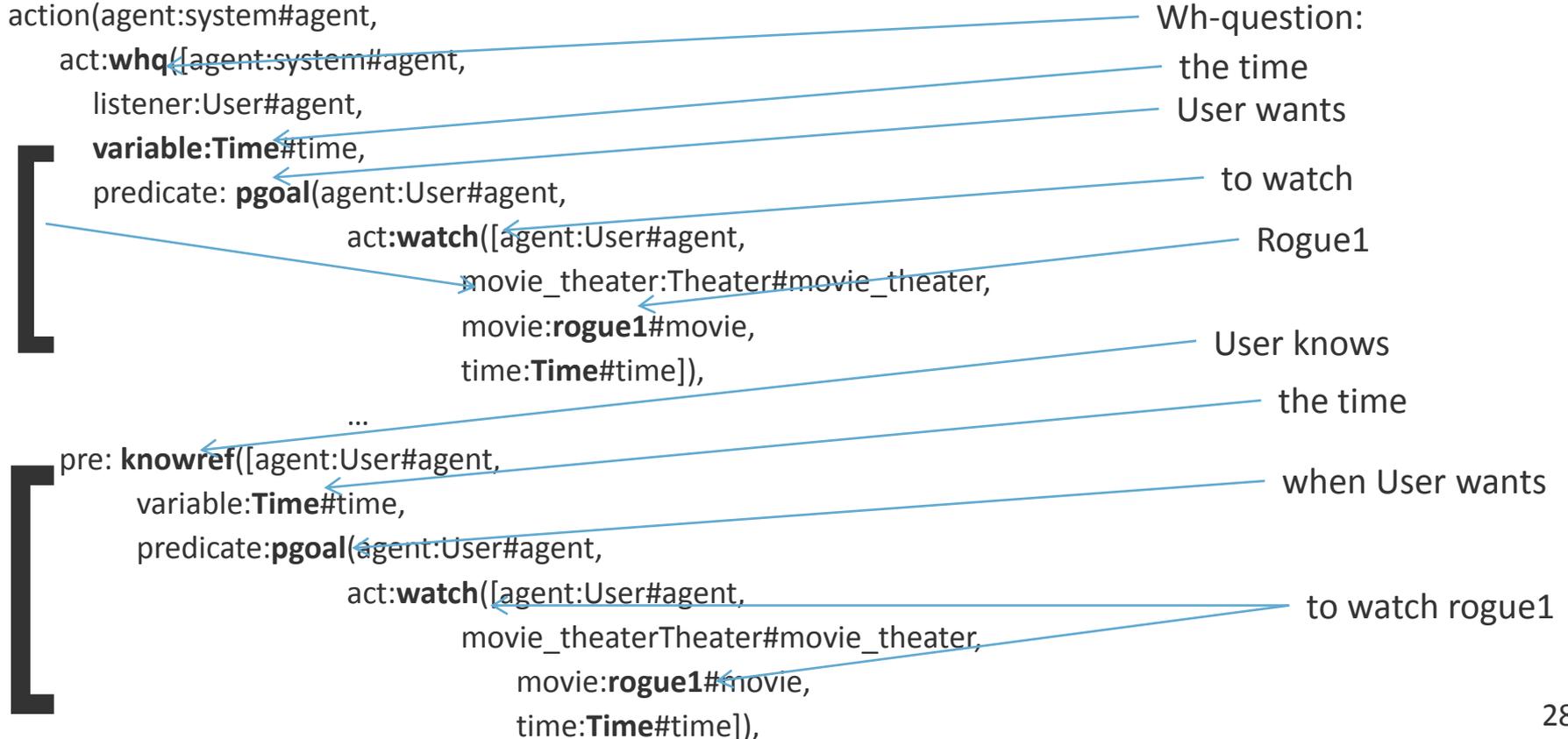
Pre: Speaker **knows** the referent of the **Var** s.t. **Pred**

Effect: Listener **knows** the referent of **Var** s.t. **Pred**

# Example

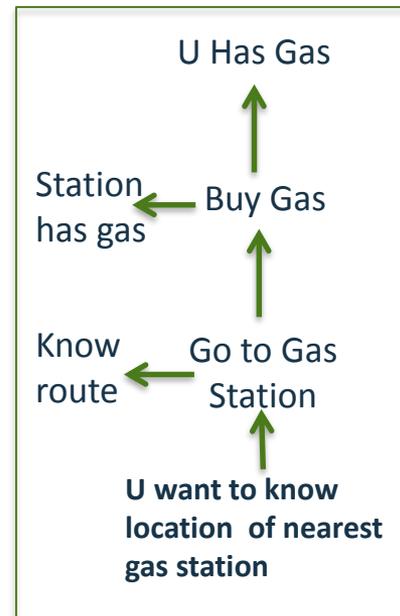
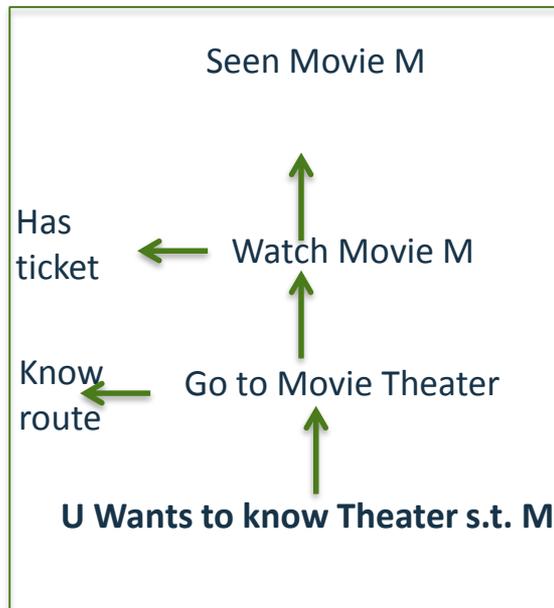
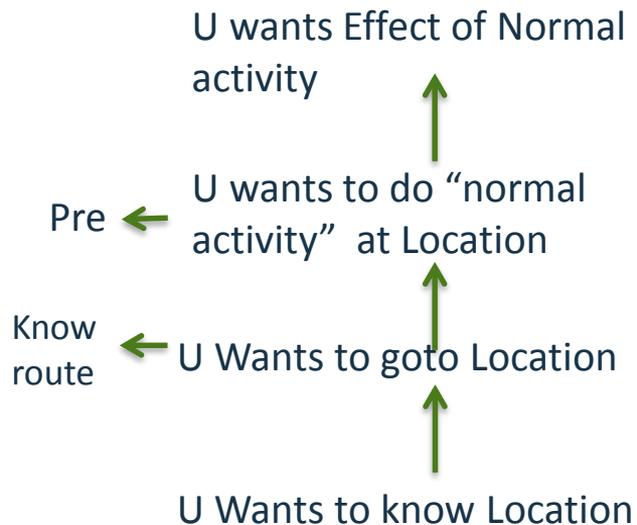
## What time do you want to watch rogue1?

© Copyright 2017 Voicebox Technologies Corporation - ICMI



# Common patterns of inference

e.g., Normal use



# Demonstrations

© Copyright 2017 Voicebox Technologies Corporation - ICMI

- ***Where is Arrival playing?***

- » The movie theater that is showing arrival at 1900 is the roxy.
- » Would you like me to purchase tickets to arrival?
- » Yes
- » How many tickets to the 1900 showing of arrival at the roxy can I purchase for you?

Wh-Q

Inform

Offer

slot-filling question

- ***Where is Dunkirk playing?***

- » The theater that is showing dunkirk at 1830 is the amc10.
- » Would you like me to purchase tickets for you for dunkirk?
- » I am very sorry that there are no tickets available for the 1830 showing of Dunkirk at amc10
- » I suggest that you watch Dunkirk on TV channel 104 at 1000.

Wh-Q

Inform

Offer,

Inform plan failure via Apology

Find alternative, Collaborative suggestion

# Demonstrations

- I want to eat at chez panisse
  - » Would you like me to reserve chez panisse for you?
  - » Yes
  - » What date would you like to eat at chez panisse?
  - » Monday
  - » How many people are in your party ?

- I am hungry

Indirect request  
Offer by means of a Y/N question

Question-asking (“Slot-filling”)  
by reasoning about K

Negative → positive state, question-asking

# Explainability

- Major issue for AI systems, especially those based on Deep Learning
- Plan-based systems are *inherently* explainable
  - » (Moore & Swartout, '91), ...
- **“Why did you say / do that?”**
- **Example:**

» **U: Where is Arrival playing?**

» **S: The movie theater that is showing Arrival at 1900 is the Roxy**

» **U: Why did you say that?**

» **S: The reason that the system informed u1 that the Roxy is showing arrival at 1900 is because system wants u1 to know the theater that is showing arrival**

## Plan

[u1,wants,to,have,tickets,to,arrival]  
*in order to achieve*  
 [u1,intends,that,u1,watch,arrival]

[u1,wants,u1,to,know,the,movie,theater,at,which,  
 u1,will,want,to,watch,arrival]  
*in order to achieve*  
 [u1,intends,that,u1,watch,arrival]

[u1#agent,wants,u1,to,know,the,time,that,  
 u1,wants,to,watch, arrival]  
*in order to achieve*  
 [u1,intends,that,u1,watch,arrival]

# Next steps

- Initial Experiment building a collaborative dialogue agent based on plan-recognition
- True planner/plan recognizer
  - » Where do the actions and their definitions come from? – E.g., crowd sourcing, build ontology
  - » How much plan recognition is needed? (wide + shallow vs narrow + deep?)
  - » Reasoning: Horn-clause modal reasoner, add: equality, constraint reasoning , abduction
  - » Negation as failure
- Hybrid Symbolic/Statistical plan-based reasoning
  - » What probabilistic reasoning will we do?
  - » Where will we get the data?
  - » Learn to Reason -- use the planner to create reasoning traces.
    - (Similar to user simulators for frame-based dialogue systems, learn from plan-based reasoning traces (uptraining?))
  - » Relationship to FSA's of speech acts – compile 1<sup>st</sup> principles reasoning into FSA's?
- Deployment issues:
  - » How to “author dialogue flows”? -- system determines speech act(s) to perform, customer determines how
  - » How to give the system a “personality” reflected in utterance / vocabulary choices
- Multimodal I/O

# Conclusions

- Current approaches to dialogue are limited
- Collaborative interaction is a key component of useful dialogue systems
- Build plan-based dialogue systems that can reason about others
- Builds on capabilities we learn at very young age, and expected by all
- Explainability via plans and meta-interpreter
- Learn to reason?