# **Dialogue and Consciousness**

Jason Baldridge (with many slides from Oliver Lemon, James Henderson, and Geert-Jan Kruijff) Department of Linguistics The University of Texas at Austin Language and Computers



#### Communication breakdown



- SNL skit: http:// www.rjauburn.com/files/ SpeechRecoDate.wmv
- Excerpt from Daniel Tammet's book "Born on a Blue Day", regarding the difficulties of having conversation when one has Asperger's.



#### The Turing Test and Loebner Prize



- Turing test: a test for machine intelligence (Alan Turing, "Computing Machinery and Intelligence", 1950)
  - engage a computer in open domain natural language conversation (through text) with a human judge who then must say whether they were talking to a real person or a computer
- Loebner Prize: Turing test competition for chatterbots with a yearly prize (\$2000) and a full Turing Test prize (\$100,000, not yet awarded).
- Previous winners of the yearly prize:
  - 2010: Suzette: seems to use a lot of topic changing
  - 2009: Do-Much-More: moves conversation from specific to general
  - 2008: Elbot [http://www.elbot.com/] (Seems like Eliza++ to me...)
  - 2007: Ultra HAL Assistant: pattern matching with some sophisticated linguistics and storing of state information
  - 2006: Jabberwacky: uses learning from previous dialogs to respond



- 2015: Rose: http://brilligunderstanding.com/ rosedemo.html
- 2012: Chip Vivant: another Eliza++
- 2010: Suzette: seems to use a lot of topic changing
- 2009: Do-Much-More: moves conversation from specific to general
- 2008: Elbot [http://www.elbot.com/] (Seems like Eliza++ to me...)
- 2007: Ultra HAL Assistant: pattern matching with some sophisticated linguistics and storing of state information
- 2006: Jabberwacky: uses learning from previous dialogs to respond



- As users talk with Jabberwacky and Cleverbot, they store their utterances and uses them in the future
- Can "learn" any language if given examples: they choose a "canned" utterance based on some simple probabilities. This is similar to spam/ham identification.
- Uses no deep representations: maps utterances to utterances
- Interestingly, this is deeply connected to how machine translation works.



- Entertainment: make robots more enjoyable, make characters in videogames more interactive
- Could collect information about people in a friendly way: talking to real people, but guide the conversation toward learning about closed domain of ages, locations, etc (automated knowledge extraction)
- Digital secretary: retrieve information about appointments, web pages, etc.
- Customer service: based on past help center conversations, provide contextually appropriate answersl
- Clearly, we have a LOOOONG way to go...
  - http://gizmodo.com/5910812/the-best-chatbot-in-the-world-is-still-awful



 Google, Facebook and many others are creating more sophisticated chatbots.

#### Helpdesk

Machine: hi

Human: hello Machine: could you please let me know what are the operating systems you are using ? Human: i am using linux Machine: ok it was in the ssl vpn on which you stated it was broken Machine: please check if you are logged in to corp network Machine: what is the error that you are running please Human: i am seeing an error related to vpn Machine: what is the error message that you are getting when connecting to vpn using network connect ?

Human: connection refused or something like that

Machine: may i know the version of network connect you connect?

#### Waxing philosophical

Human: what is the purpose of life ? Machine: to serve the greater good . Human: what is the purpose of living ? Machine: to live forever . Human: what is the purpose of existence ? Machine: to find out what happens when we get to the planet earth . Human: where are you now ? Machine: i 'm in the middle of nowhere . Human: what is the purpose of dying ? Machine: to have a life . Human: what is the purpose of being intelligent ? Machine: to find out what it is .

Machine: to find out what it is . Human: what is the purpose of emotions? Machine: i don 't know.

Vinyals and Lee (2015) "A Neural Conversational Model": http://arxiv.org/pdf/1506.05869v1.pdf



Markus Dickinson's dialogue slides



- The next set of slides are from a course taught by Oliver Lemon and James Henderson.
- Original Edinburgh slides available from Oliver Lemon's website:
  - http://homepages.inf.ed.ac.uk/olemon/lecture1.pdf

LACK 13

## What is a Dialogue System?

- An interface that supports natural language input, responds using natural language, and emulates (some) human conversational skills:
  - Recognising dialogue structures
  - Contextual interpretation
  - Turn taking
  - Managing conversational obligations
  - Response planning



See http://www.sigdial.org/

August 2006



## Vision for end-users

Natural dialogue communication requires :

- Content: users say what they want
- Flexibility: users say it the way they want to, with no need to learn the specific commands a device or service "expects"
- Adaptivity: system adapts to preferences, knowledge, and ability of user, and to context
- Learning: systems use dialogue strategies learned from experience, and continue to learn from interactions with users



» Automatic optimisation



23

August 2006

#### LALK 24

## Task Complexity

## Application dependent



Simple

## Directly affects:

- Dialogue complexity
- Types and quantity of system knowledge

Complexity of system's reasoning abilities





August 2006

## **Spoken Dialogue Systems**



LALK

14

### WITAS demo: <a href="http://www-csli.stanford.edu/semlab-hold/witas/">http://www-csli.stanford.edu/semlab-hold/witas/</a>







#### • Ford and Microsoft: SYNC for in-car entertainment

- hands-free calling
- voice control of MP3 player
  - "play artist Rolling Stones", "play track 'Sour Times'
- text-to-speech: reads incoming text messages
- English, Spanish, and Canadian French

CMU Communicator, flight reservation system with limited domain TTS <a href="http://festvox.org/ldom/ldom\_com.html">http://festvox.org/ldom/ldom\_com.html</a>

000	S: Hello. Welcome to the Carnegie Mellon Communicator. Are you a registered user? U: "HI THIS IS ALEX RUDNICKY I'D LIKE TO FLY TO CHARLESTON PLEASE"
001	S: Hello, dr Rudnicky, i'm glad to hear from you again. You can ask me for help at any time. You can fly to charleston in south carolina or charleston in west virginia. U: "SOUTH CAROLINA"
002	S: a flight to charleston. departing pittsburgh on what day U: "ON THE FIFTEENTH "
003	S(1): I'm sorry. I must have misheard the date a flight on Tuesday, February fifteenth. okay what day would you be traveling? U: "ON MARCH FIFTEENTH"
004	S(1): traveling on Wednesday, March fifteenth. S(2): One Moment Please Let's see flights from pittsburgh to charleston if you'll just hold on, this shouldn't take too long Gosh, the network has been slow all day I'm glad you called, though 
005	S: the flight after that departs Pittsburgh at two ten p.m. arriving in Charleston at four ten p.m. Is that OK? U: "OKAY THAT'S FINE"



- Telephone based dialog systems connect the user to a database of information.
- The WITAS demo included a visual map that was being discussed, but the agent was still disembodied.
- Opportunities and challenges arise when we put a face and/or body to the voice!

An interlude...

# Dialog systems with robots: slides from CoSy project by Geert-Jan Kruijff



- We'd like to be able to talk with robots
  - ... it is a natural way of interacting with social agents
  - ... it is an intuitive means of coordination & control, even in teams
- Unfortunately, robots don't always see the world like we do
  - To have a meaningful conversation with a robot, it needs to *understand*
  - But understanding goes well beyond *linguistic* understanding!
  - We talk about the world, and doing things there
  - The big issue is thus, how we can relate what we say, to an understanding of what we can do where, and why, and how; i.e. to *situated understanding*

### How robots tend to see reality



Deutsches Forschungszentnum für Künstliche Intelligenz Germen Besaarch Center für Artificial Intelligenze





© 2005 Geert-Jan M. Kruijff

University of Texas @ Austin, March 22 2006



- Legos! (Who said we can't play. Serious)
  - Fixed dialogue sequence(s): FSAs model dialogue
  - Limited domain speech recognition
  - Very limited sensor input





### "Slightly" more complex 'bots



Deutsches Forschungszentnum für Künstliche Intelligenz





http://www.dfki.de/cosy/www/

#### What does he know?



Deutsches Forschungszentnum für Künstliche Intelligenz. German Besearch Genter für Artificial Intelligenze





### Sure he knows ...



Deutsches Forschungszentnum für Künstliche Intelligenz German Research Center für Artificial Intelligenzen





## Spoken language analysis





### There are still a few issues ...

- How do we link language to "reality"?
  - What different aspects of "reality" do we link to?
  - What is in a link? (coordination; cross-modal association; hierarchical processing)
  - What does a link affect? (expectations, attention; spatiotemporal interpretability)
- What can we learn from all this?
  - Understanding language means more than understanding language
  - What does that mean for how language reflects reality, and affects our perception, understanding of reality?
  - What does that mean for how language is learnt?
- What does human-robot interaction look like?









- These robots are expert systems designed to work with a particular domain
- Unlike an expert system that works entirely in the computer (e.g. interfacing to a database), robots must ground their conceptual representations to real objects: (some portion of) the real world is the database.
  - Expressions like "the door" must be mapped to specific door (e.g. DOOR\_42) that is an object with certain boundaries at a certain location in the robot's map.
- Dialog systems for robots thus must ground their concepts in reality and also know how linguistic expressions refer to those concepts.



- It is important to map linguistic expressions (like words) to a formal representation, but this is not understanding.
- We need to connect those representations to things in our domain (as on the previous slide), and furthermore be able to connect them with our current understanding.
- A famous system from the 70's called SHRDLU provides an excellent view into this.



- Dialog system for interacting with a blocks world
- Circumscribed domain: easy to encode its current state
  - red, blue, and green things
  - pyramids and cubes
  - relative position of object

















- We have seen a number of shallow methods:
  - statistical text classification based on bags of words
  - machine translation alignment based on bags of words
  - ELIZA, Jabberwacky, and other chatbots
- And deep methods:
  - SHRDLU
  - The CoSy robot



- Presumably humans perform a fairly deep analysis. Should we be trying to model how humans process language?
- Deep methods may not reflect how humans do it, but they derive in spirit from it: attempt to represent meaning abstractly, connect it to the world, and reason about it
- Shallow methods play on strengths of computers: fast pattern matching, accurate counting, access to much larger volumes of data



- Think about transportation: how to get from point A to point B?
  - Fly: like a bird? like a plane? some other way?
  - Travel virtually
- How we solve the problem depends on the needs
  - information exchange, physical contact
- Also depends on the resources:
  - fuel costs, internet infrastructure



- Work until the 80's was dominated by deep approaches
- Statistical methods based on shallow analyses became dominant in the 90's.
- Now deep and shallow methods are coming back together as the statistical methods have become more sophisticated and efficient (along with increase in cheap computing power)
- Almost all work now is on a continuum between shallow and deep, often mixing deep and shallow methods.
- Most are building airplanes, not birds: creating models that do not necessarily bear a direct relationship with how humans process language.



- A "deep" syntactic analysis can still be seen as quite shallow: we may translate a sentence into a formal representation, but rarely is that connected to actual things in the world (they are not grounded).
  - So, a meaning representation is just that: a representation -- it doesn't "mean" anything unless it is connected to a model of the world
- The robots you've seen are good examples of depth achieved in a (very) limited domain.
  - They hear a sentence, turn it into a formal representation, connect the symbols to things in the world, and take action
  - But: are they thinking?



- What does it mean to think, or to be intelligent, or be conscious? Is it enough to manipulate symbols and produce the right kind of response and behaviors?
- Somehow, most people feel it isn't, and that consciousness lies at the heart of the issue.
- This issue is a part of the question of monism versus dualism: are mind and body the same kind of thing or are they different?
  - material monism: only material things exist; consciousness emerges from material processes and structures
  - **dualism**: both material and mental/spiritual realms exist; consciousness is generated from the mental realm



- The idea of a body without a mind is an old one.
- Golems were soulless beings created from inanimate matter (e.g. clay) by men.
  - golem: "cocoon", modern Hebrew, derived from gelem "raw material"
- Interestingly, a common feature of golems in stories is that they lacked the ability to speak.

- Actually, the robots are "just" doing symbol manipulation -- is that really thinking?
- John Searle argued that these sorts of applications of computational "intelligence" in fact aren't intelligent at all, using the Chinese Room
- A man who doesn't speak Chinese sits in a room and responds to messages in Chinese by using a rule book to transform those into (Chinese) responses.
- Does the man "know" Chinese? How about the room?





• The argument attacks what is termed "strong AI":

The appropriately programmed computer with the right inputs and outputs would thereby have a mind in exactly the same sense human beings have minds.

- Searle's thought experiment can be pulled apart, attacked, and defended from various angles. He claims that the Chinese room can (in theory) pass the Turing test: Searle argues that it is not conscious of what it is doing and is thus not truly *understanding* its inputs and outputs.
- Much ink has been spilled on this topic: see the summary on Wikipedia for some highlights: <u>http://en.wikipedia.org/wiki/Chinese\_room</u>



- Systems reply: the whole room understands Chinese, including the rulebooks, pencils, etc.
- **Robot reply**: embed the Chinese room in a robot that interacts with the physical world, thus providing it with a causal connection to its inputs and outputs. The man in the room is trapped, manipulating symbols that are meaningless to him, yet they are grounded for the robot via its sensors.
- Brain simulator reply (redesigning the room): write a program that simulates the action of every neuron in a Chinese speaker's brain. (Or replace each neuron of a speaker, one at a time. Does he lose consciousness? If so, at what point?)



- Other responses have focused not on suggesting how the Chinese room actually does or could have a mind, but instead on raising doubts about the very nature of the thought experiment and its reliance on our intuition.
- **Speed and complexity**: the room is overly simplistic and could never pass muster because responses would take far too long and would require storage of astronomical proportions.
- Other minds: it is hard to tell whether other people are actually thinking! (I.e., the Chinese room "proves" that humans don't have minds.)
  - Daniel Dennett: natural selection would prefer zombies (people who act just like us, but don't actually have minds, and are thus simpler)



- So, where does that leave us? Somehow, most people seem to have an intuition that there needs to be some sense in which a sentient being is conscious of their thought processes -- that they are not just following a system of rules that were pre-specified, e.g., by some programmer. Is there a ghost in the machine, a homunculus?
  - Homunculus: literally, means "little man" and refers to scale models of human beings in general. In consciousness studies, it refers to the idea that there is a person within a body who is perceiving and interpreting input from the body's various senses.
- Homunculi have an obvious problem: how do they perceive and interpret the input they receive?
- An alternative is to consider consciousness as an emergent property of many small and simple, unaware computational units, i.e. neurons.

#### Emergence and the computational theory of mind

- Radiolab: The Unconscious Toscanini of the Brain:
- From wikipedia: <u>http://en.wikipedia.org/wiki/Emergence</u>

http://www.wnyc.org/shows/radiolab/episodes/2005/02/18



Image of visual cortex neurons from: http://8ight.stumbleupon.com/tag/neuroscience/



Homunculus, from wikipedia: http://en.wikipedia.org/wiki/Homunculus

46

- From wikipedia: <u>http://en.wikipedia.org/wiki/Emergence</u>:
  - **Emergence**: An emergent behaviour or emergent property can appear when a number of simple entities (agents) operate in an environment, forming more complex behaviours as a collective.
- Recall the Radiolab discussion of coffee, in which the recognition of coffee by the brian involves many individual, connected neurons firing "smell", "color", "taste".
- Neural networks are one way of simulating such effects, and are used for many practical pattern recognition tasks.

## A simple neural network





- Qualia
  - Still, there is a sense in which we "experience" things very directly, *including* our thought processes and how we feel about them. These senses are defined as qualia: "the way things seem to us"
  - David Chalmers: distinguishes "easy" and "hard" problems of consciousness studies:
    - The easy problems are those of finding neural mechanisms and explaining cognitive functions: the ability to discriminate and categorize environmental stimuli, the capacity to verbally report mental states... The hard problem is that of *experience*: why does all this processing give rise to an experienced inner life at all?
  - Again, this may very well be tackled as an emergent property of brain activity.





- Back to talking, possibly thinking machines... What about the question of programming behaviors and whether the machine actually is aware?
- Does it make a difference is we instead set a bunch of initial conditions and then let the robot learn from experience?
  - Think back to rule-based versus statistically-based spam filters? What representations did they use? What interactions did they assume?
- Is thinking meat fundamentally different from thinking machines?



"They're made out of meat."

"Meat?"

```
"Meat. They're made out of meat."
```

"Meat?"

"There's no doubt about it. We picked up several from different parts of the planet, took them aboard our recon vessels, and probed them all the way through. They're completely meat."

"That's impossible. What about the radio signals? The messages to the stars?"

"They use the radio waves to talk, but the signals don't come from them. The signals come from machines."

"So who made the machines? That's who we want to contact."

"They made the machines. That's what I'm trying to tell you. Meat made the machines."

"That's ridiculous. How can meat make a machine? You're asking me to believe in sentient meat."

"I'm not asking you, I'm telling you. These creatures are the only sentient race in that sector and they're made out of meat."

"Maybe they're like the orfolei. You know, a carbon-based intelligence that goes through a meat stage."

(continued -- see the web link above)



- We are still in the dark on our own mental processes.
  - Were we programmed?
  - Are we actually aware of how we work?
  - Can we ever figure out just how we tick?
  - How objectively do we interpret the real world?
- Colorblind people definitely see things differently. There are more extreme varieties of mental experience, such as synesthesia, a condition in which information and perceptions from different senses overlap, such as associating colors and textures with letters and numbers.
  - Listen in class to excerpt from Daniel Tammet's *Born on a Blue Day* regarding the perception of numbers for a person with synesthesia.

http://en.wikipedia.org/wiki/File:Synesthesia.svg



- The biologists are already engineering new life. Artificial intelligence is still pretty much unintelligent, in the general sense. But what if we could actually do it?
  - What if they don't play nice? (Terminator, etc.)
  - What if they are smarter than us? (Stanislaw Lem, Imaginary Magnitude, Golem XIV)
  - What rights would they have? (Bladerunner)
  - Would they be moral individuals? (What does that mean?)
  - Can we encode ourselves in digital form, and possibly embed our mental experience in other bodies? (Avatar, Surrogates)

#### Are we just "building stuff"?



 Listen to segment from Radiolab's "(So-called) Life" episode: genetic engineering making bacteria smell

better: http://www.wnyc.org/shows/radiolab/episodes/2008/03/14



 The implications of what is done with science and technology reach far and wide, in all areas of life.