Intelligence and Physical Symbol Systems

Classic AI and the Chinese Room

Carlotta Pavese

11.12.13

Outline

Outline

Review

The Computational Theory of the Mind

The Computational Theory of the Mind (CTM)

Minds are an organization of representations, or, more precisely, an organizer of representations.

Review

The Computational Theory of the Mind

The Computational Theory of the Mind (CTM)

Minds are an organization of representations, or, more precisely, an organizer of representations.

Functionalism

"Minds are what brains do." (Minsky, *Society of Mind*)

Review

The Computational Theory of the Mind

The Computational Theory of the Mind (CTM)

Minds are an organization of representations, or, more precisely, an organizer of representations.

Functionalism

"Minds are what brains do." (Minsky, *Society of Mind*)

The Computational Brain

Brains organize representations, which is another way of saying that they compute.

Review

Representation

Representations

Computational representations are patterns of bits, i.e. a binary signal or sequences thereof.

But, representations aren't just any bit patterns:

Review

Representation

Representations

Computational representations are patterns of bits, i.e. a binary signal or sequences thereof.

But, representations aren't just any bit patterns:

Representation

These bit patterns stand for something, they represent:

- They covary with an external thing
- They get used by the computational system in a way that exploits this covariance

Physical Symbol Systems

The Definition

Physical Symbol System PSS (Newell & Simon)

1. **Symbols**: contains a set of interpretable and combinable items (also called *representations*)

Physical Symbol Systems

The Definition

Physical Symbol System PSS (Newell & Simon)

1. **Symbols**: contains a set of interpretable and combinable items (also called *representations*)
   - These bit patterns covary with something, call it x, outside the system

Physical Symbol Systems

The Definition

Physical Symbol System PSS (Newell & Simon)

1. **Symbols**: contains a set of interpretable and combinable items (also called *representations*)
   - These bit patterns covary with something, call it x, outside the system
   - This covariance allows the system to behave in a way that depends on x

Physical Symbol Systems

The Definition

Physical Symbol System PSS (Newell & Simon)

1. **Symbols**: contains a set of interpretable and combinable items (also called *representations*)
   - These bit patterns covary with something, call it x, outside the system
   - This covariance allows the system to behave in a way that depends on x
2. **Operations**: the interpretation and combination of symbols can be broken down into a set of more basic processes (read, write, copy)

Physical Symbol Systems

The Definition

Physical Symbol System PSS (Newell & Simon)

1. **Symbols**: contains a set of interpretable and combinable items (also called *representations*)
   - These bit patterns covary with something, call it x, outside the system
   - This covariance allows the system to behave in a way that depends on x
2. **Operations**: the interpretation and combination of symbols can be broken down into a set of more basic processes (read, write, copy)

- PSS: a computer whose bit patterns are representations

Physical Symbol Systems

The Hypothesis

Physical Symbol System Hypothesis (Newell & Simon)

Physical symbol systems have the necessary and sufficient means for intelligent action.

- We've seen some reasons, for going along with this, let's review these

Review

Making CTM Look Inevitable

1. Thinking is moving from one thought to another

Review

Making CTM Look Inevitable

1. Thinking is moving from one thought to another
   - In a way that preserves truth

Review

Making CTM Look Inevitable

1. Thinking is moving from one thought to another
   - In a way that preserves truth
2. Thoughts are representations

Review

Making CTM Look Inevitable

1. Thinking is moving from one thought to another
   - In a way that preserves truth
2. Thoughts are representations
   - These are physical things

Review

Making CTM Look Inevitable

1. Thinking is moving from one thought to another
   - In a way that preserves truth
2. Thoughts are representations
   - These are physical things
3. Formal logic discovered rules shuffling representations while preserving truth

Review

Making CTM Look Inevitable

1. Thinking is moving from one thought to another
   - In a way that preserves truth
2. Thoughts are representations
   - These are physical things
3. Formal logic discovered rules shuffling representations while preserving truth
4. This is what computers do! Without a homunculus!

Review

Making CTM Look Inevitable

1. Thinking is moving from one thought to another
   - In a way that preserves truth
2. Thoughts are representations
   - These are physical things
3. Formal logic discovered rules shuffling representations while preserving truth
4. This is what computers do! Without a homunculus!
5. Folk psychology? (beliefs and desires cause actions)

Review

Making CTM Look Inevitable

1. Thinking is moving from one thought to another
   - In a way that preserves truth
2. Thoughts are representations
   - These are physical things
3. Formal logic discovered rules shuffling representations while preserving truth
4. This is what computers do! Without a homunculus!
5. Folk psychology? (beliefs and desires cause actions)
   - Computation is a causal process involving representations; beliefs and desires are representations

Review

Making CTM Look Inevitable

1. Thinking is moving from one thought to another
   - In a way that preserves truth
2. Thoughts are representations
   - These are physical things
3. Formal logic discovered rules shuffling representations while preserving truth
4. This is what computers do! Without a homunculus!
5. Folk psychology? (beliefs and desires cause actions)
   - Computation is a causal process involving representations; beliefs and desires are representations
6. Fodor: so thinking is probably just computing!

Review

Turing's Test

- Turing proposed to replace *Can a computer think?* with *Is the computer indistinguishable from a human in conversation?*

Review

Turing's Test

- Turing proposed to replace *Can a computer think?* with *Is the computer indistinguishable from a human in conversation?*
- This requires that the computer be capable of:

Review

Turing's Test

- Turing proposed to replace *Can a computer think?* with *Is the computer indistinguishable from a human in conversation?*
- This requires that the computer be capable of:
  - Strategic reasoning

Review

Turing's Test

- Turing proposed to replace *Can a computer think?* with *Is the computer indistinguishable from a human in conversation?*
- This requires that the computer be capable of:
  - Strategic reasoning
  - Language use

Review

Turing's Test

- Turing proposed to replace *Can a computer think?* with *Is the computer indistinguishable from a human in conversation?*
- This requires that the computer be capable of:
  - Strategic reasoning
  - Language use
- Language use makes this task really hard:

Review

Turing's Test

- Turing proposed to replace *Can a computer think?* with *Is the computer indistinguishable from a human in conversation?*
- This requires that the computer be capable of:
  - Strategic reasoning
  - Language use
- Language use makes this task really hard:
  - Language is infinite: combinatorial explosion

Review

Turing's Test

- Turing proposed to replace *Can a computer think?* with *Is the computer indistinguishable from a human in conversation?*
- This requires that the computer be capable of:
  - Strategic reasoning
  - Language use
- Language use makes this task really hard:
  - Language is infinite: combinatorial explosion
  - Requires combining knowledge of the world with grammatical rules

Review

Turing's Test

- Turing proposed to replace *Can a computer think?* with *Is the computer indistinguishable from a human in conversation?*
- This requires that the computer be capable of:
  - Strategic reasoning
  - Language use
- Language use makes this task really hard:
  - Language is infinite: combinatorial explosion
  - Requires combining knowledge of the world with grammatical rules
- So it seems that there's something to the Turing Test

Review

Turing's Test

- Turing proposed to replace *Can a computer think?* with *Is the computer indistinguishable from a human in conversation?*
- This requires that the computer be capable of:
  - Strategic reasoning
  - Language use
- Language use makes this task really hard:
  - Language is infinite: combinatorial explosion
  - Requires combining knowledge of the world with grammatical rules
- So it seems that there's something to the Turing Test
- Computers are well-positioned to at least handle combinatorial explosion

Where We Are

With the Computational Theory

- A computational approach nicely explains the combinatorial features of language

Where We Are

With the Computational Theory

- A computational approach nicely explains the combinatorial features of language
- But combining world knowledge and grammatical knowledge in a human-like way is still on the frontier of computational linguistics

Where We Are

With the Computational Theory

- A computational approach nicely explains the combinatorial features of language
- But combining world knowledge and grammatical knowledge in a human-like way is still on the frontier of computational linguistics
- More generally, building machines that understand what words mean is on the frontier of research

Where We Are

With the Computational Theory

- A computational approach nicely explains the combinatorial features of language
- But combining world knowledge and grammatical knowledge in a human-like way is still on the frontier of computational linguistics
- More generally, building machines that understand what words mean is on the frontier of research
- But what exactly are we talking about when we talk about the meaning of a word or thought?

Where We Are

With the Computational Theory

- A computational approach nicely explains the combinatorial features of language
- But combining world knowledge and grammatical knowledge in a human-like way is still on the frontier of computational linguistics
- More generally, building machines that understand what words mean is on the frontier of research
- But what exactly are we talking about when we talk about the meaning of a word or thought?
- This where philosophers are useful (maybe...)

Outline

The Chinese Room

First version

The Chinese Room

1. Suppose someone is in a room with a complete manual that tells one how to answer any Chinese question in Chinese.

The Chinese Room

First version

The Chinese Room

1. Suppose someone is in a room with a complete manual that tells one how to answer any Chinese question in Chinese.
2. The person cannot understand Chinese.

The Chinese Room

First version

The Chinese Room

1. Suppose someone is in a room with a complete manual that tells one how to answer any Chinese question in Chinese.
2. The person cannot understand Chinese.
3. From one door, she gets a sheet of paper with a question formulated in Chinese.

The Chinese Room

First version

The Chinese Room

1. Suppose someone is in a room with a complete manual that tells one how to answer any Chinese question in Chinese.
2. The person cannot understand Chinese.
3. From one door, she gets a sheet of paper with a question formulated in Chinese.
4. She follows the manual, and outputs a correct response in Chinese to the opposite door.

The Chinese Room

First version

The Chinese Room

1. Suppose someone is in a room with a complete manual that tells one how to answer any Chinese question in Chinese.
2. The person cannot understand Chinese.
3. From one door, she gets a sheet of paper with a question formulated in Chinese.
4. She follows the manual, and outputs a correct response in Chinese to the opposite door.
5. She does not thereby count as understanding Chinese.

The Chinese Room

The Argument

The Chinese Room

1. Passing the Turing test is not sufficient for intelligence.

The Chinese Room

The Argument

The Chinese Room

1. Passing the Turing test is not sufficient for intelligence.
2. Intelligence behavior requires understanding.

The Chinese Room

The Argument

The Chinese Room

1. Passing the Turing test is not sufficient for intelligence.
2. Intelligence behavior requires understanding.
3. But one could pass Turing test (in Chinese or English), without understanding any word of Chinese or English.

The Chinese Room

The Argument

The Chinese Room

1. Passing the Turing test is not sufficient for intelligence.
2. Intelligence behavior requires understanding.
3. But one could pass Turing test (in Chinese or English), without understanding any word of Chinese or English.
4. Hence,Turing test is not sufficient to understanding.

The Chinese Room

Objection to the set up

The Chinese Room

1. No person in such a situation will pass the Turing test!

The Chinese Room

Objection to the set up

The Chinese Room

1. No person in such a situation will pass the Turing test!
2. No instruction manual will be enough complete!

The Chinese Room

Second version

The Chinese Room

1. Suppose someone has come up with a program P that passes the Turing Test in Chinese when run on a digital computer

The Chinese Room

Second version

The Chinese Room

1. Suppose someone has come up with a program P that passes the Turing Test in Chinese when run on a digital computer
2. Now, put Searle, who doesn't understand Chinese, in a room with an 'input slot', P, baskets of tiles w/Chinese characters on them & an 'output slot'

The Chinese Room

Second version

The Chinese Room

1. Suppose someone has come up with a program P that passes the Turing Test in Chinese when run on a digital computer
2. Now, put Searle, who doesn't understand Chinese, in a room with an 'input slot', P, baskets of tiles w/Chinese characters on them & an 'output slot'
3. By following P, Searle could fool a Chinese speaker into thinking that they are communicating with another Chinese speaker

The Chinese Room

Second version

The Chinese Room

1. Suppose someone has come up with a program P that passes the Turing Test in Chinese when run on a digital computer
2. Now, put Searle, who doesn't understand Chinese, in a room with an 'input slot', P, baskets of tiles w/Chinese characters on them & an 'output slot'
3. By following P, Searle could fool a Chinese speaker into thinking that they are communicating with another Chinese speaker
4. **Searle** (p.18): I still don't understand Chinese!

The Chinese Room

Second version

The Chinese Room

1. Suppose someone has come up with a program P that passes the Turing Test in Chinese when run on a digital computer
2. Now, put Searle, who doesn't understand Chinese, in a room with an 'input slot', P, baskets of tiles w/Chinese characters on them & an 'output slot'
3. By following P, Searle could fool a Chinese speaker into thinking that they are communicating with another Chinese speaker
4. **Searle** (p.18): I still don't understand Chinese!
5. There's no important difference between a digital computer & the Chinese Room, so despite passing the Turing Test, a digital computer could never actually understand language

The Chinese Room

The Intuition Behind it All

- Searle's real contention is that to understand a natural language you have to know what the words mean, i.e. their semantics

The Chinese Room

The Intuition Behind it All

- Searle's real contention is that to understand a natural language you have to know what the words mean, i.e. their semantics
- But since computers are 'purely formal' they are only really sensitive to syntax

The Chinese Room

The Intuition Behind it All

- Searle's real contention is that to understand a natural language you have to know what the words mean, i.e. their semantics
- But since computers are 'purely formal' they are only really sensitive to syntax
- So a computer could never *understand* language

The Chinese Room

The Intuition Behind it All

- Searle's real contention is that to understand a natural language you have to know what the words mean, i.e. their semantics
- But since computers are 'purely formal' they are only really sensitive to syntax
- So a computer could never *understand* language
- It could never possess intentionality

The Chinese Room

The Intuition Behind it All

- Searle's real contention is that to understand a natural language you have to know what the words mean, i.e. their semantics
- But since computers are 'purely formal' they are only really sensitive to syntax
- So a computer could never *understand* language
- It could never possess intentionality
  - *Intentionality* is *aboutness* (Franz Brentano)

The Chinese Room

The Intuition Behind it All

- Searle's real contention is that to understand a natural language you have to know what the words mean, i.e. their semantics
- But since computers are 'purely formal' they are only really sensitive to syntax
- So a computer could never *understand* language
- It could never possess intentionality
  - *Intentionality* is *aboutness* (Franz Brentano)
  - Human thoughts and words are not just symbols, they are about something!

Outline

The Chinese Room

Response 1: the Systems Reply

1. In the Chinese Room, Searle is the FSM

The Chinese Room

Response 1: the Systems Reply

1. In the Chinese Room, Searle is the FSM
2. CTM does not claim that the FSM understands the representations

The Chinese Room

Response 1: the Systems Reply

1. In the Chinese Room, Searle is the FSM
2. CTM does not claim that the FSM understands the representations
   - Remember: no homunculus!

The Chinese Room

Response 1: the Systems Reply

1. In the Chinese Room, Searle is the FSM
2. CTM does not claim that the FSM understands the representations
   - Remember: no homunculus!
3. It claims that the whole implemented program counts as an implementation of understanding

The Chinese Room

Response 1: the Systems Reply

1. In the Chinese Room, Searle is the FSM
2. CTM does not claim that the FSM understands the representations
   - Remember: no homunculus!
3. It claims that the whole implemented program counts as an implementation of understanding
4. Does the whole room consist of a system that understands Chinese?

Outline

The Chinese Room

Response 2: is the room a PSS?

- Is the Chinese Room actually a PSS?

The Chinese Room

Response 2: is the room a PSS?

- Is the Chinese Room actually a PSS?
  - Are the symbols actually *representations*?

The Chinese Room

Response 2: is the room a PSS?

- Is the Chinese Room actually a PSS?
  - Are the symbols actually *representations*?
  - Physical symbol systems aren't just any old computer

The Chinese Room

Response 2: is the room a PSS?

- Is the Chinese Room actually a PSS?
  - Are the symbols actually *representations*?
  - Physical symbol systems aren't just any old computer
  - Its symbols must depend on the external world and be used in a way that exploits this dependence

The Chinese Room

Response 2: is the room a PSS?

- Is the Chinese Room actually a PSS?
  - Are the symbols actually *representations*?
  - Physical symbol systems aren't just any old computer
  - Its symbols must depend on the external world and be used in a way that exploits this dependence

The Big Question

What exactly do we need to find in a bit pattern to count it as representing something?

Outline

Stanley

Response 2.1 The Brain simulator reply

- These critics concede Searle's claim that just running a natural language processing program as described in the CR scenario does not create any understanding, whether by a human or a computer system.

Stanley

Response 2.1 The Brain simulator reply

- These critics concede Searle's claim that just running a natural language processing program as described in the CR scenario does not create any understanding, whether by a human or a computer system.
- But these critics hold that a variation on the computer system could understand.

Stanley

Response 2.1 The Brain simulator reply

- These critics concede Searle's claim that just running a natural language processing program as described in the CR scenario does not create any understanding, whether by a human or a computer system.
- But these critics hold that a variation on the computer system could understand.
- It might be a system that simulated the detailed operation of an entire brain, neuron by neuron (The Brain Simulator Reply).

Stanley

Response 2.1 The Brain simulator reply

- Consider a computer that operates in quite a different manner than the usual AI program with scripts and operations on strings of linguistic symbols.

Stanley

Response 2.1 The Brain simulator reply

- Consider a computer that operates in quite a different manner than the usual AI program with scripts and operations on strings of linguistic symbols.
- The Brain Simulator reply asks us to suppose instead the program simulates the actual sequence of nerve firings that occur in the brain of a native Chinese language speaker when that person understands Chinese—every nerve, every firing.

Stanley

Response 2.1 The Brain simulator reply

- Consider a computer that operates in quite a different manner than the usual AI program with scripts and operations on strings of linguistic symbols.
- The Brain Simulator reply asks us to suppose instead the program simulates the actual sequence of nerve firings that occur in the brain of a native Chinese language speaker when that person understands Chinese—every nerve, every firing.
- Since the computer then works the very same way as the brain of a native Chinese speaker, processing information in just the same way, it will understand Chinese.

Outline

Stanley

Response 2.2 The Robot reply

- The Robot Reply concedes Searle is right about the Chinese Room scenario.

Stanley

Response 2.2 The Robot reply

- The Robot Reply concedes Searle is right about the Chinese Room scenario.
- it shows that a computer trapped in a computer room cannot understand language, or know what words mean.

Stanley

Response 2.2 The Robot reply

- The Robot Reply concedes Searle is right about the Chinese Room scenario.
- it shows that a computer trapped in a computer room cannot understand language, or know what words mean.
- the Robot Reply suggests that we put a digital computer in a robot body, with sensors, such as video cameras and microphones, and add effectors, such as wheels to move around with, and arms with which to manipulate things in the world.

Stanley

Response 2.2 The Robot reply

- The Robot Reply concedes Searle is right about the Chinese Room scenario.
- it shows that a computer trapped in a computer room cannot understand language, or know what words mean.
- the Robot Reply suggests that we put a digital computer in a robot body, with sensors, such as video cameras and microphones, and add effectors, such as wheels to move around with, and arms with which to manipulate things in the world.
- Such a robot—a computer with a body—could do what a child does, learn by seeing and doing.

Outline

The Chinese Room

Response 3: Are our intuitions reliable?

- Should we really listen to our intuitions about what counts as understanding?

The Chinese Room

Response 3: Are our intuitions reliable?

- Should we really listen to our intuitions about what counts as understanding?
  - Are our intuitions reliable?

The Chinese Room

Response 3: Are our intuitions reliable?

- Should we really listen to our intuitions about what counts as understanding?
  - Are our intuitions reliable?
  - If a computer passes the Turing test, would not that trump our pre-theoretical intuitions?

The Chinese Room

Response 3: Are our intuitions reliable?

- Should we really listen to our intuitions about what counts as understanding?
  - Are our intuitions reliable?
  - If a computer passes the Turing test, would not that trump our pre-theoretical intuitions?
  - Why should we give our intuitions a role in science?

The Chinese Room

What is understanding?

- What is understanding anyway?

The Chinese Room

What is understanding?

- What is understanding anyway?
  - It is not enough to say that a computer does not understand.

The Chinese Room

What is understanding?

- What is understanding anyway?
  - It is not enough to say that a computer does not understand.
  - We should say what is required of understanding.

The Chinese Room

What is understanding?

- What is understanding anyway?
  - It is not enough to say that a computer does not understand.
  - We should say what is required of understanding.
  - But if one shows to be able to manipulate meaningful symbols, what more is required for one to count as understanding?