Consciousness II

Mechanical Consciousness

Carlotta Pavese

11.19.13

Outline

Outline

What is Consciousness?

Some Things are Not So Baffling or Hard to Explain

Access Consciousness

Conscious Experience

What is Consciousness?

Some Things are Not So Baffling or Hard to Explain

Access Consciousness

- Awareness, ability to discriminate and react to stimuli

Conscious Experience

What is Consciousness?

Some Things are Not So Baffling or Hard to Explain

Access Consciousness

- Awareness, ability to discriminate and react to stimuli
- Ability to focus attention, deliberately change behavior, access own mental states, etc.

Conscious Experience

What is Consciousness?

Some Things are Not So Baffling or Hard to Explain

Access Consciousness

- Awareness, ability to discriminate and react to stimuli
- Ability to focus attention, deliberately change behavior, access own mental states, etc.

Conscious Experience

- What it's is like when you are perceiving, experiencing and thinking

What is Consciousness?

Some Things are Not So Baffling or Hard to Explain

Access Consciousness

- Awareness, ability to discriminate and react to stimuli
- Ability to focus attention, deliberately change behavior, access own mental states, etc.

Conscious Experience

- What it's is like when you are perceiving, experiencing and thinking
- The subjective aspect of consciousness

The Study of Mind
Takes on Consciousness

Chalmers and Others

- Functionalism can't explain conscious experience

The Study of Mind
Takes on Consciousness

Chalmers and Others

- Functionalism can't explain conscious experience
- Focus on a particular mental activity, specify its function and the mechanism that performs it

The Study of Mind
Takes on Consciousness

Chalmers and Others

- Functionalism can't explain conscious experience
- Focus on a particular mental activity, specify its function and the mechanism that performs it
  - What's left?

Dennett and Others

- There's nothing left once you've specified functions and mechanisms
- That there's something left is an illusion

The Study of Mind
Takes on Consciousness

Chalmers and Others

- Functionalism can't explain conscious experience
- Focus on a particular mental activity, specify its function and the mechanism that performs it
  - What's left?
  - What it's like to perform that mental activity

Dennett and Others

- There's nothing left once you've specified functions and mechanisms
- That there's something left is an illusion

The Study of Mind
Takes on Consciousness

Chalmers and Others

- Functionalism can't explain conscious experience
- Focus on a particular mental activity, specify its function and the mechanism that performs it
  - What's left?
  - What it's like to perform that mental activity
- Functionalism needs to be supplemented

Dennett and Others

- There's nothing left once you've specified functions and mechanisms
- That there's something left is an illusion

The Hard Problem

Dramatizing it with Jackson's Thought Experiment

Jackson's Thought Experiment: Mary and B&W room

The Hard Problem

Dramatizing it with Chalmers' Zombies

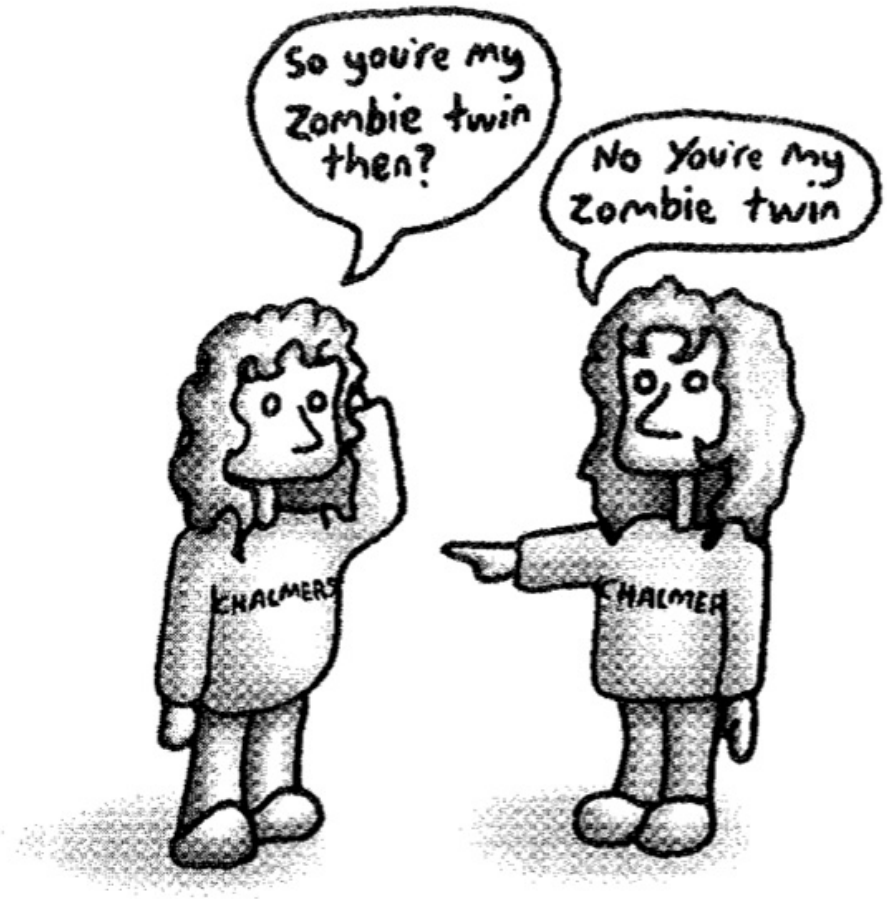


FIGURE 2.4 • Which is which? Can you tell? Can they?

The Hard Problem

Dramatizing it with Chalmers' Zombies

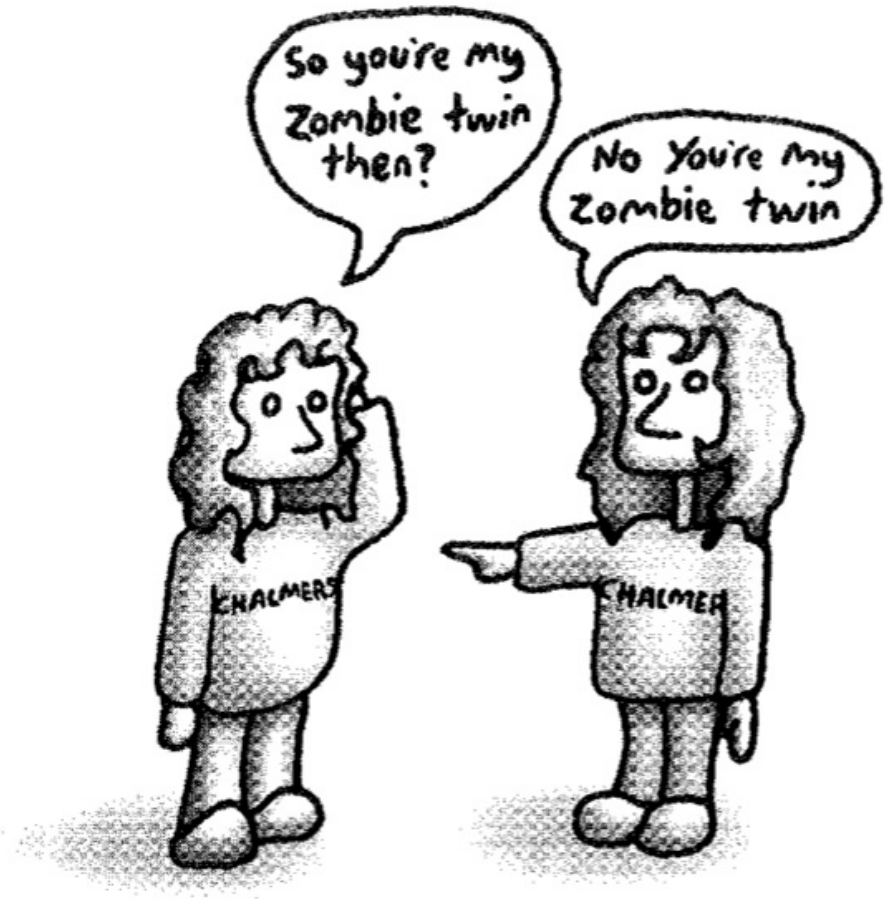


FIGURE 2.4 • Which is which? Can you tell? Can they?

- "A zombie is just something physically identical to me, but which has no conscious experience – all is dark inside" (Chalmers *The Conscious Mind* p.96)

The Hard Problem

Dramatizing it with Chalmers' Zombies

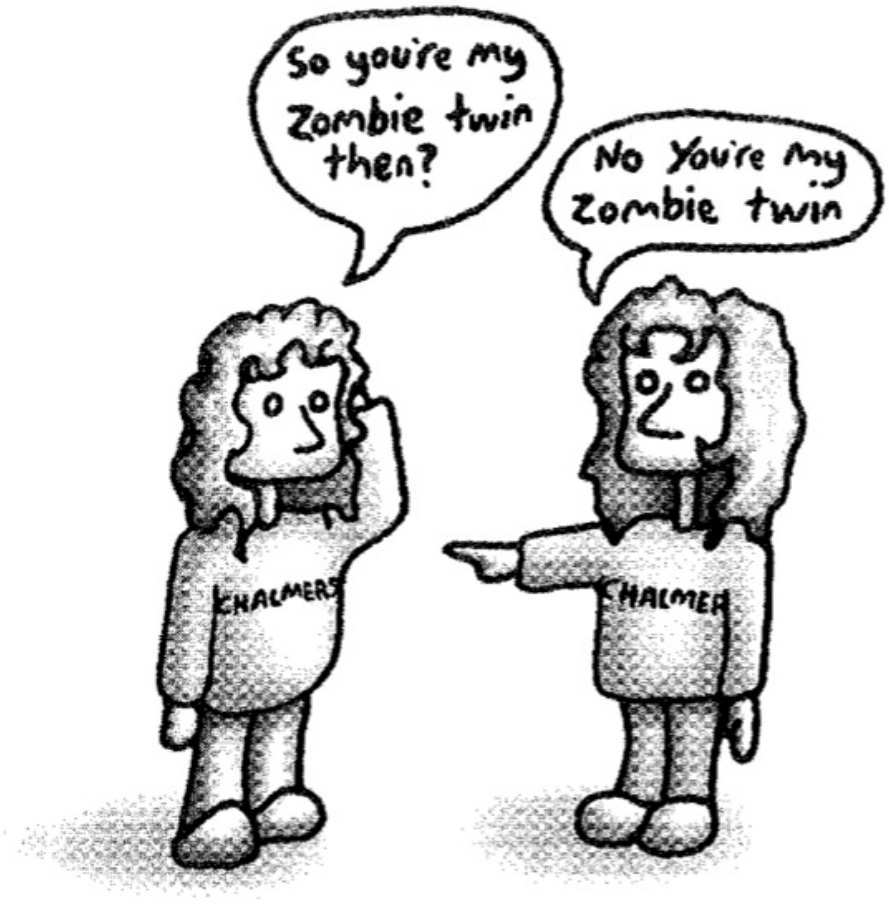


FIGURE 2.4 • Which is which? Can you tell? Can they?

- "A zombie is just something physically identical to me, but which has no conscious experience – all is dark inside" (Chalmers *The Conscious Mind* p.96)
- That zombies are conceivable illustrates the conceptual gap between experience and causal organization

The Consciousness Debate
What Do the Views Have Going for Them?

Chalmers et. al. (Anti-Reductionists)

Dennett et. al. (Reductionists)

The Consciousness Debate
What Do the Views Have Going for Them?

Chalmers et. al. (Anti-Reductionists)

- Respects intuitions in thought experiments and reflection that consciousness is 'something extra'

Dennett et. al. (Reductionists)

The Consciousness Debate

What Do the Views Have Going for Them?

Chalmers et. al. (Anti-Reductionists)

- Respects intuitions in thought experiments and reflection that consciousness is 'something extra'
- Respects idea that we really can never know what it is like to be a bat

Dennett et. al. (Reductionists)

The Consciousness Debate
What Do the Views Have Going for Them?

Chalmers et. al. (Anti-Reductionists)

- Respects intuitions in thought experiments and reflection that consciousness is 'something extra'
- Respects idea that we really can never know what it is like to be a bat

Dennett et. al. (Reductionists)

- Doesn't require revising our theory of the universe

The Consciousness Debate
What Do the Views Have Going for Them?

Chalmers et. al. (Anti-Reductionists)

- Respects intuitions in thought experiments and reflection that consciousness is 'something extra'
- Respects idea that we really can never know what it is like to be a bat

Dennett et. al. (Reductionists)

- Doesn't require revising our theory of the universe
- Doesn't require new of laws and science

The Consciousness Debate
What Do the Views Need to Do Better?

Chalmers et. al. (Anti-Reductionists)

- Say precisely what it is about conscious experience that is resistant to functional & neurophysiological analysis

The Consciousness Debate
What Do the Views Need to Do Better?

Chalmers et. al. (Anti-Reductionists)

- Say precisely what it is about conscious experience that is resistant to functional & neurophysiological analysis
  - Self, perspectival, agentive, feeling, ineffable?

The Consciousness Debate
What Do the Views Need to Do Better?

Chalmers et. al. (Anti-Reductionists)

- Say precisely what it is about conscious experience that is resistant to functional & neurophysiological analysis
  - Self, perspectival, agentive, feeling, ineffable?
- Make it palatable to have experiences be fundamental constituents of reality

The Consciousness Debate
What Do the Views Need to Do Better?

Chalmers et. al. (Anti-Reductionists)

- Say precisely what it is about conscious experience that is resistant to functional & neurophysiological analysis
  - Self, perspectival, agentive, feeling, ineffable?
- Make it palatable to have experiences be fundamental constituents of reality

Dennett et. al. (Reductionists)

- Address systematically why we seem so tempted to see conscious experience as 'something extra'

The Consciousness Debate
What Do the Views Need to Do Better?

Chalmers et. al. (Anti-Reductionists)

- Say precisely what it is about conscious experience that is resistant to functional & neurophysiological analysis
  - Self, perspectival, agentive, feeling, ineffable?
- Make it palatable to have experiences be fundamental constituents of reality

Dennett et. al. (Reductionists)

- Address systematically why we seem so tempted to see conscious experience as 'something extra'
- Actually explain the distinctive features of conscious experience in terms of functions and mechanisms

The Consciousness Debate

The Reductionist Strategy: doing better than Dennett

Self (Damasio, Metzinger, et. al.)

How is it that experiences seem to belong to and, over time, partly constitute *you*? What function does this 'binding' serve?

The Consciousness Debate

The Reductionist Strategy: doing better than Dennett

Self (Damasio, Metzinger, et. al.)

How is it that experiences seem to belong to and, over time, partly constitute *you*? What function does this 'binding' serve?

Perspective (Marr, Edelman et. al.)

What's the geometry of our visual field? What's distinctive about the touches, sights, sounds and smells from this vantage point? How is this reflected in computational and neurophysiological accounts?

The Consciousness Debate

The Reductionist Strategy: doing better than Dennett

Self (Damasio, Metzinger, et. al.)

How is it that experiences seem to belong to and, over time, partly constitute *you*? What function does this 'binding' serve?

Perspective (Marr, Edelman et. al.)

What's the geometry of our visual field? What's distinctive about the touches, sights, sounds and smells from this vantage point? How is this reflected in computational and neurophysiological accounts?

Agentive (Franklin, Damasio, et. al.)

How does this sense of control and choice over my thoughts and movements come about? What role does it play?

The Consciousness Debate

The Reductionist Strategy: doing better than Dennett

Feeling

Why do experiences have a feeling at all? How do those feelings differ and what role to they play in the system?

The Consciousness Debate

The Reductionist Strategy: doing better than Dennett

Feeling

Why do experiences have a feeling at all? How do those feelings differ and what role to they play in the system?

Ineffability

Is your experience is ineffable? Or is it just impossible in practice to communicate every detail?

Mechanical Consciousness
Reductionist Option

Artificial Consciousness (Reductionist)

If the reductionist strategy succeeds will it be, in principal, possible to produce artificial consciousness?

- However, our kind of consciousness might be radically dependent on our mechanisms/embodiment

Mechanical Consciousness
Reductionist Option

Artificial Consciousness (Reductionist)

If the reductionist strategy succeeds will it be, in principal, possible to produce artificial consciousness?

- Yes: it's all functions or mechanisms

- However, our kind of consciousness might be radically dependent on our mechanisms/embodiment
  - Eye-wiring, placement of ear drums, etc.

Mechanical Consciousness
Reductionist Option

Artificial Consciousness (Reductionist)

If the reductionist strategy succeeds will it be, in principal, possible to produce artificial consciousness?

- Yes: it's all functions or mechanisms

- However, our kind of consciousness might be radically dependent on our mechanisms/embodiment
  - Eye-wiring, placement of ear drums, etc.
  - Parallel to cricket phonotaxis

Mechanical Consciousness
Reductionist Option

Artificial Consciousness (Reductionist)

If the reductionist strategy succeeds will it be, in principal, possible to produce artificial consciousness?

- Yes: it's all functions or mechanisms

- However, our kind of consciousness might be radically dependent on our mechanisms/embodiment
  - Eye-wiring, placement of ear drums, etc.
  - Parallel to cricket phonotaxis
- Artificial consciousness may require artificial humans

Mechanical Consciousness
Reductionist Option

Artificial Consciousness (Reductionist)

If the reductionist strategy succeeds will it be, in principal, possible to produce artificial consciousness?

- Yes: it's all functions or mechanisms

- However, our kind of consciousness might be radically dependent on our mechanisms/embodiment
  - Eye-wiring, placement of ear drums, etc.
  - Parallel to cricket phonotaxis
- Artificial consciousness may require artificial humans
- Anyway, we don't even know what functions and mechanisms to implement so it's a LONG way off.

Outline

Outline

Mechanical Consciousness

Anti-reductionist Option

Artificial Consciousness (Anti-reductionist)

"A non-reductive view of consciousness does not automatically lead to a pessimistic view of AI, however.

Mechanical Consciousness

Anti-reductionist Option

Artificial Consciousness (Anti-reductionist)

"A non-reductive view of consciousness does not automatically lead to a pessimistic view of AI, however. The two issues are quite separate.

Mechanical Consciousness

Anti-reductionist Option

Artificial Consciousness (Anti-reductionist)

"A non-reductive view of consciousness does not automatically lead to a pessimistic view of AI, however. The two issues are quite separate. The first concerns the *strength* of the connection between physical systems and consciousness: is consciousness constituted by physical processes or does it merely arise from physical processes?

Mechanical Consciousness

Anti-reductionist Option

Artificial Consciousness (Anti-reductionist)

"A non-reductive view of consciousness does not automatically lead to a pessimistic view of AI, however. The two issues are quite separate. The first concerns the *strength* of the connection between physical systems and consciousness: is consciousness constituted by physical processes or does it merely arise from physical processes? The second concerns the *shape* of the connection: just *which* physical systems give rise to consciousness?" (Chalmers 1996, p.314)

Focusing on the Second Question

What is the relation between functional organization and experience?

- The functional organization of a mind:

Focusing on the Second Question

What is the relation between functional organization and experience?

- The functional organization of a mind:
  - What causes what; it's causal dynamics!

Focusing on the Second Question

What is the relation between functional organization and experience?

- The functional organization of a mind:
  - What causes what; it's causal dynamics!
- Suppose consciousness is something more than just a functional organization

Focusing on the Second Question

What is the relation between functional organization and experience?

- The functional organization of a mind:
  - What causes what; it's causal dynamics!
- Suppose consciousness is something more than just a functional organization
  - Still: what is the relation between functional organization and conscious experience?

Focusing on the Second Question

What is the relation between functional organization and experience?

- The functional organization of a mind:
  - What causes what; it's causal dynamics!
- Suppose consciousness is something more than just a functional organization
  - Still: what is the relation between functional organization and conscious experience?

Chalmers' Organizational Invariance Principle

Any two systems with the same fine-grained functional organization will have qualitatively identical experiences.

Artificial Consciousness

It's Even Possible for the Anti-reductionist

Chalmers' Organizational Invariance Principle

Any two systems with the same fine-grained functional organization will have qualitatively identical experiences.

- The functional organization of human minds is a computational system

Artificial Consciousness

It's Even Possible for the Anti-reductionist

Chalmers' Organizational Invariance Principle

Any two systems with the same fine-grained functional organization will have qualitatively identical experiences.

- The functional organization of human minds is a computational system
- So a computer can match your functional organization

Artificial Consciousness

It's Even Possible for the Anti-reductionist

Chalmers' Organizational Invariance Principle

Any two systems with the same fine-grained functional organization will have qualitatively identical experiences.

- The functional organization of human minds is a computational system
- So a computer can match your functional organization
- By this principle, it will have identical experiences and will thus be conscious!

Artificial Consciousness

It's Even Possible for the Anti-reductionist

Chalmers' Organizational Invariance Principle

Any two systems with the same fine-grained functional organization will have qualitatively identical experiences.

- The functional organization of human minds is a computational system
- So a computer can match your functional organization
- By this principle, it will have identical experiences and will thus be conscious!
- What's the argument for this principle!?

Chalmers' Thought Experiment

An Argument for the Invariance Principle

- Suppose the principle is false, so there are two identically (functionally) organized systems w/ different experiences

Chalmers' Thought Experiment

An Argument for the Invariance Principle

- Suppose the principle is false, so there are two identically (functionally) organized systems w/ different experiences
- Let one be silicon and the other neurons

Chalmers' Thought Experiment

An Argument for the Invariance Principle

- Suppose the principle is false, so there are two identically (functionally) organized systems w/ different experiences
- Let one be silicon and the other neurons
- We can gradually transform the latter into the former by replacing neurons with chips

Chalmers' Thought Experiment

An Argument for the Invariance Principle

- Suppose the principle is false, so there are two identically (functionally) organized systems w/ different experiences
- Let one be silicon and the other neurons
- We can gradually transform the latter into the former by replacing neurons with chips
- Somewhere along the line, the experience must become different; call this neuron N

Chalmers' Thought Experiment

An Argument for the Invariance Principle

- Suppose the principle is false, so there are two identically (functionally) organized systems w/ different experiences
- Let one be silicon and the other neurons
- We can gradually transform the latter into the former by replacing neurons with chips
- Somewhere along the line, the experience must become different; call this neuron N
- Take a functionally identical chip, install it in the chip/neuron network alongside N and put a switch between it and N

Chalmers' Thought Experiment

An Argument for the Invariance Principle

- Suppose the principle is false, so there are two identically (functionally) organized systems w/ different experiences
- Let one be silicon and the other neurons
- We can gradually transform the latter into the former by replacing neurons with chips
- Somewhere along the line, the experience must become different; call this neuron N
- Take a functionally identical chip, install it in the chip/neuron network alongside N and put a switch between it and N
- As this switch is flipped, the experience must change

Chalmers' Thought Experiment

An Argument for the Invariance Principle: dancing qualia

- So, as the switch is flipped, the experience changes

Chalmers' Thought Experiment

An Argument for the Invariance Principle: dancing qualia

- So, as the switch is flipped, the experience changes
- Problem: there's no way for the system to notice

Chalmers' Thought Experiment

An Argument for the Invariance Principle: dancing qualia

- So, as the switch is flipped, the experience changes
- Problem: there's no way for the system to notice
- The causal organization stays constant, so there's no room for the thought 'Hmmm! Something strange just happened!'

Chalmers' Thought Experiment

An Argument for the Invariance Principle: dancing qualia

- So, as the switch is flipped, the experience changes
- Problem: there's no way for the system to notice
- The causal organization stays constant, so there's no room for the thought 'Hmmm! Something strange just happened!'
- Noticing would amount to there being some difference in downstream processing

Chalmers' Thought Experiment

An Argument for the Invariance Principle: dancing qualia

- So, as the switch is flipped, the experience changes
- Problem: there's no way for the system to notice
- The causal organization stays constant, so there's no room for the thought 'Hmmm! Something strange just happened!'
- Noticing would amount to there being some difference in downstream processing
  - But that is a function difference and there isn't one

Chalmers' Thought Experiment

An Argument for the Invariance Principle: dancing qualia

- So, as the switch is flipped, the experience changes
- Problem: there's no way for the system to notice
- The causal organization stays constant, so there's no room for the thought 'Hmmm! Something strange just happened!'
- Noticing would amount to there being some difference in downstream processing
  - But that is a function difference and there isn't one
- "This I take to be a *reductio* of the original assumption" (Chalmers 2010, p.24)

Chalmers' Thought Experiment

An Argument for the Invariance Principle: dancing qualia

- So, as the switch is flipped, the experience changes
- Problem: there's no way for the system to notice
- The causal organization stays constant, so there's no room for the thought 'Hmmm! Something strange just happened!'
- Noticing would amount to there being some difference in downstream processing
  - But that is a function difference and there isn't one
- "This I take to be a *reductio* of the original assumption" (Chalmers 2010, p.24)
- Thus the principle must be right!

Artificial Consciousness
For Anti-Reductionists

- Same kind of argument can 'transform' any natural consciousness into artificial consciousness

Artificial Consciousness
For Anti-Reductionists

- Same kind of argument can 'transform' any natural consciousness into artificial consciousness
  - Replace neurons w/functionally identical silicon chips

Artificial Consciousness
For Anti-Reductionists

- Same kind of argument can 'transform' any natural consciousness into artificial consciousness
  - Replace neurons w/functionally identical silicon chips
- And so it would seem that conscious AI should be, in principle, possible, even if Anti-reductionism is true!

Artificial Consciousness
For Anti-Reductionists

- Same kind of argument can 'transform' any natural consciousness into artificial consciousness
  - Replace neurons w/functionally identical silicon chips
- And so it would seem that conscious AI should be, in principle, possible, even if Anti-reductionism is true!
- Principle of Organizational Invariance does remain controversial among philosophers of mind who believe that 'inverted qualia' are possible (Block, Shoemaker)

Outline

Creating Artificial Consciousness

How Would we Know?

- Because conscious experience is about first-person experience, it is hard to design a test for it

Creating Artificial Consciousness

How Would we Know?

- Because conscious experience is about first-person experience, it is hard to design a test for it
- It's hard to say how we'd ever know we succeeded!

Creating Artificial Consciousness

How Would we Know?

- Because conscious experience is about first-person experience, it is hard to design a test for it
- It's hard to say how we'd ever know we succeeded!

Reflect on Trurl (Lem's "Seventh Sally")

Trurl creates a miniature, but functionally identical replica of a tyrant's kingdom for the king to abuse. By functionally identical, I mean it! By outsourcing the tyrant's cruelty, Trurl has saved him and his fellow citizens some suffering, but was it ethical?

Creating Artificial Consciousness

How Would we Know?

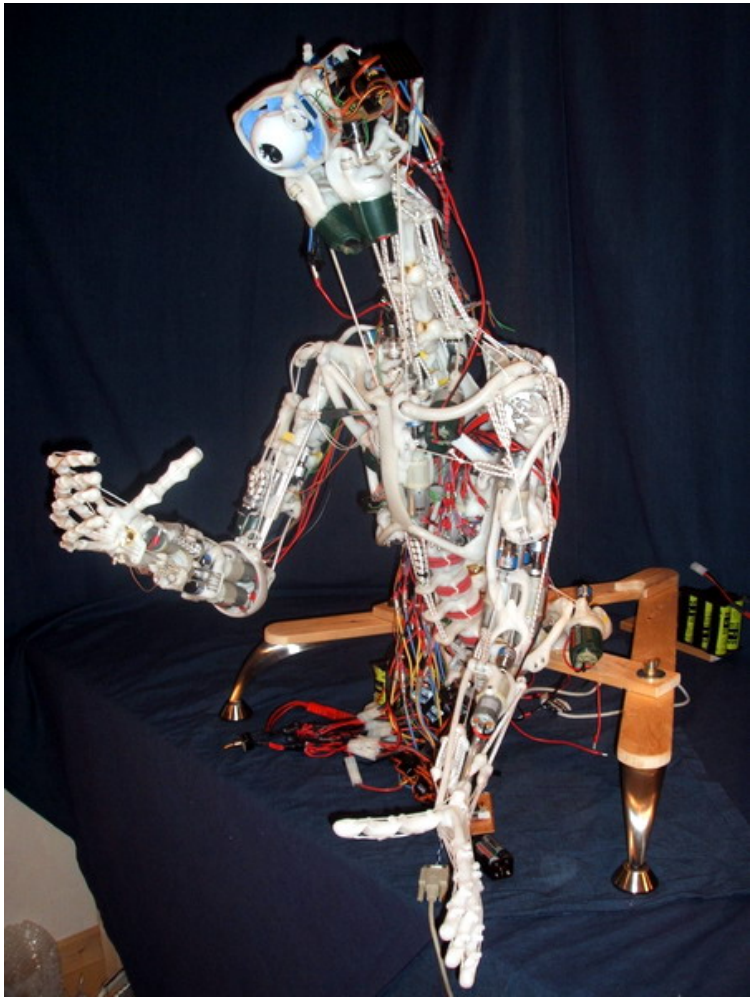
- Because conscious experience is about first-person experience, it is hard to design a test for it
- It's hard to say how we'd ever know we succeeded!

Reflect on Trurl (Lem's "Seventh Sally")

Trurl creates a miniature, but functionally identical replica of a tyrant's kingdom for the king to abuse. By functionally identical, I mean it! By outsourcing the tyrant's cruelty, Trurl has saved him and his fellow citizens some suffering, but was it ethical? Are the replicas suffering?

Artificial Consciousness

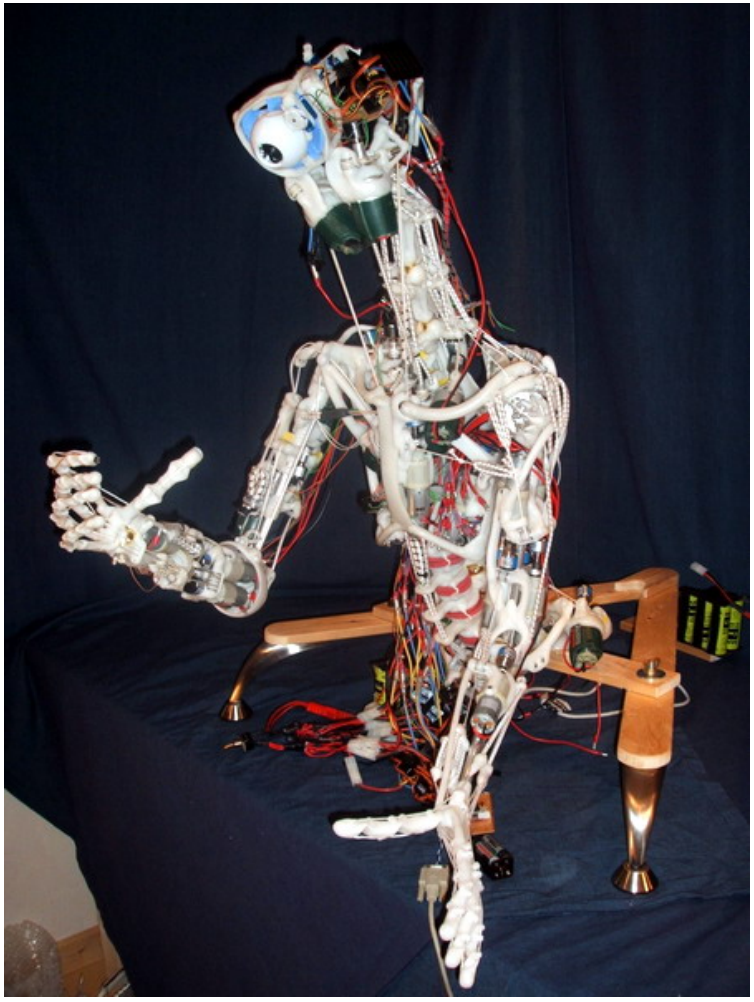
CRONOS: representation and access consciousness



- Has an internal representation of the external world

Artificial Consciousness

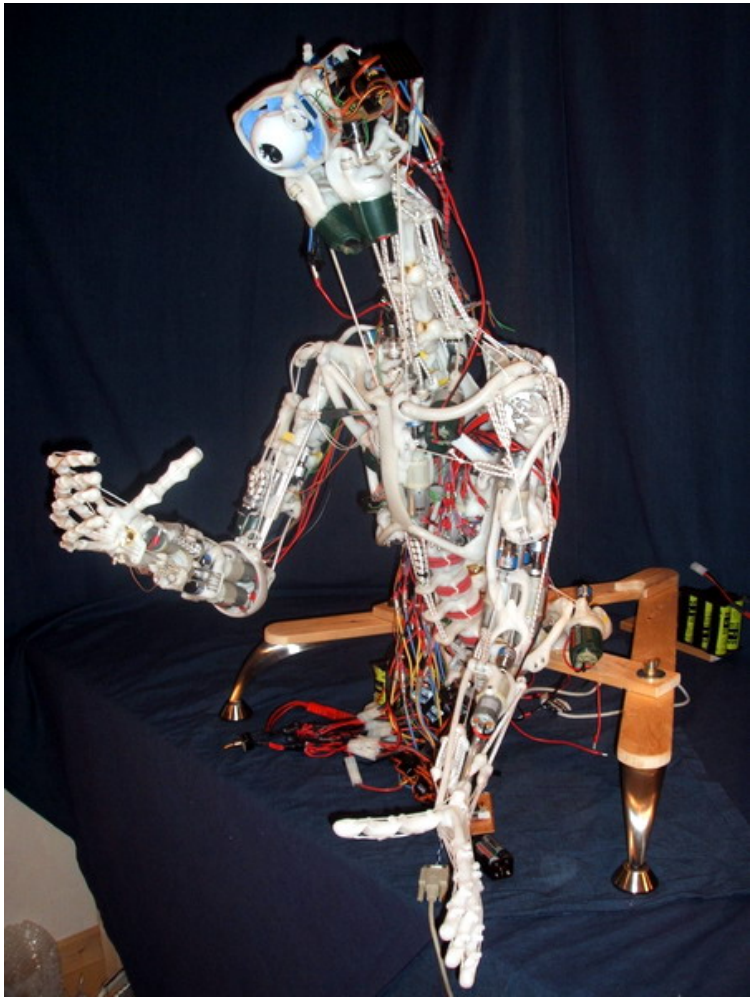
CRONOS: representation and access consciousness



- Has an internal representation of the external world
- Also: representation of itself!

Artificial Consciousness

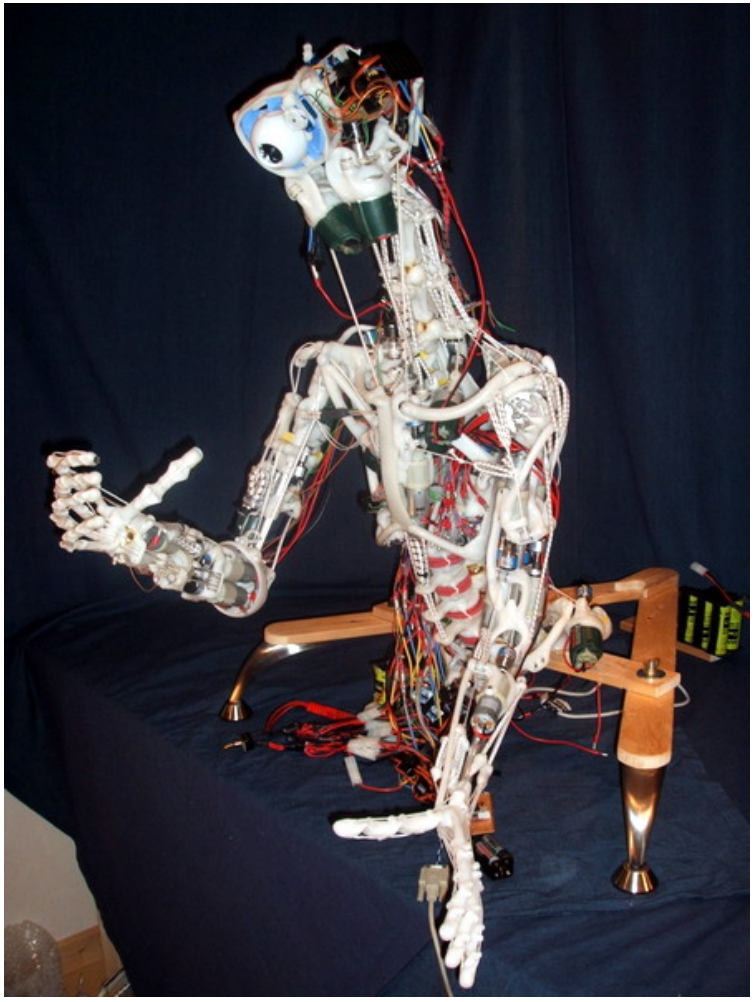
CRONOS: representation and access consciousness



- Has an internal representation of the external world
- Also: representation of itself!
- Including a representation of itself representing the world

Artificial Consciousness

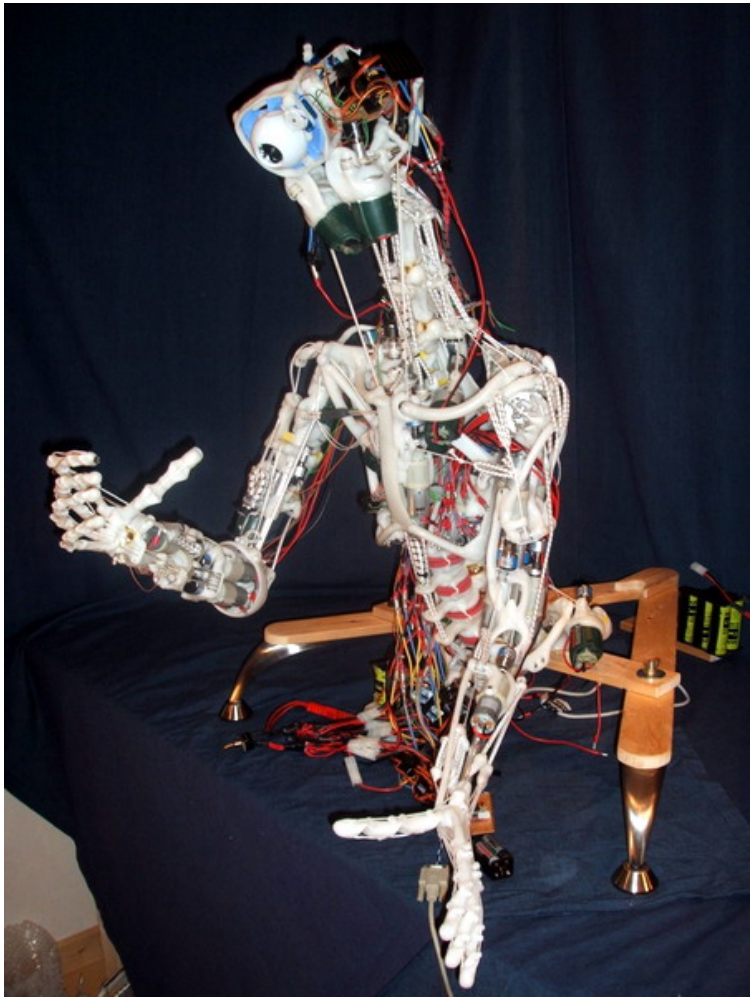
CRONOS: representation and access consciousness



- Has an internal representation of the external world
- Also: representation of itself!
- Including a representation of itself representing the world
- Model of itself captures some of what we have in mind by a 'self' and being 'self aware'

Artificial Consciousness

CRONOS: representation and access consciousness



- Has an internal representation of the external world
- Also: representation of itself!
- Including a representation of itself representing the world
- Model of itself captures some of what we have in mind by a 'self' and being 'self aware'
- `http://youtu.be/nvZjvGOgtbY`