Understanding and Modelling Consciousness

David Gamez

Middlesex University, London, UK

Consciousness, Models and the Artificial Workshop, Creativity 2019, 13th December 2019

Talk Overview

- What is consciousness?
- Thought experiments and imagination.
- Science of consciousness.
- Models of consciousness.
- Conclusion.

What is Consciousness?

Physical World?

- Every day we are immersed in a world of colourful, noisy, moving things.
- Most of the time we interpret what we see as the actual physical world.
- This is a natural and obvious!
- Position known as **naïve realism**.

Physical World?

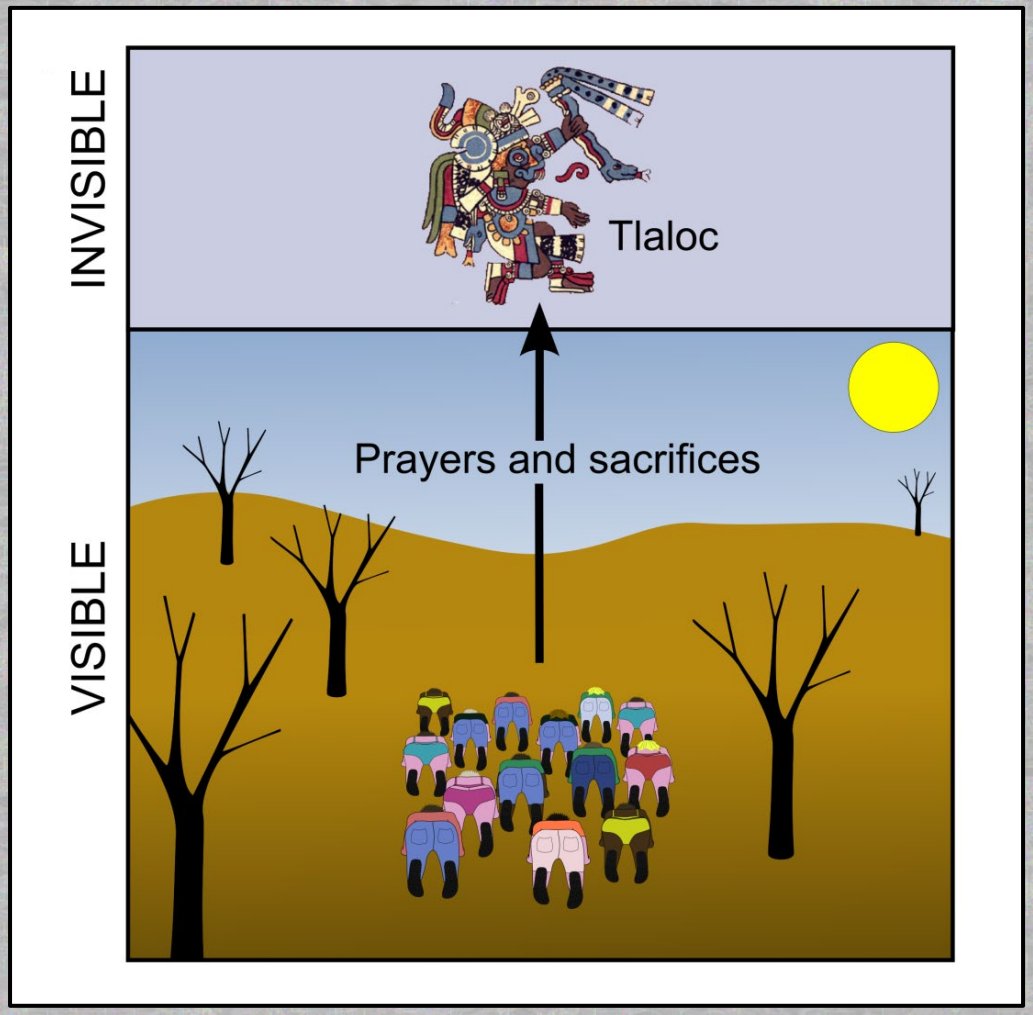
Naïve Realism

- The everyday world that we experience has *regularities*.
- If I throw a cat through the air, its colour and sound move together; its rate of acceleration can be calculated with a simple equation.
- We use things that we can't see to explain these regularities.
- These are **invisible explanations**.

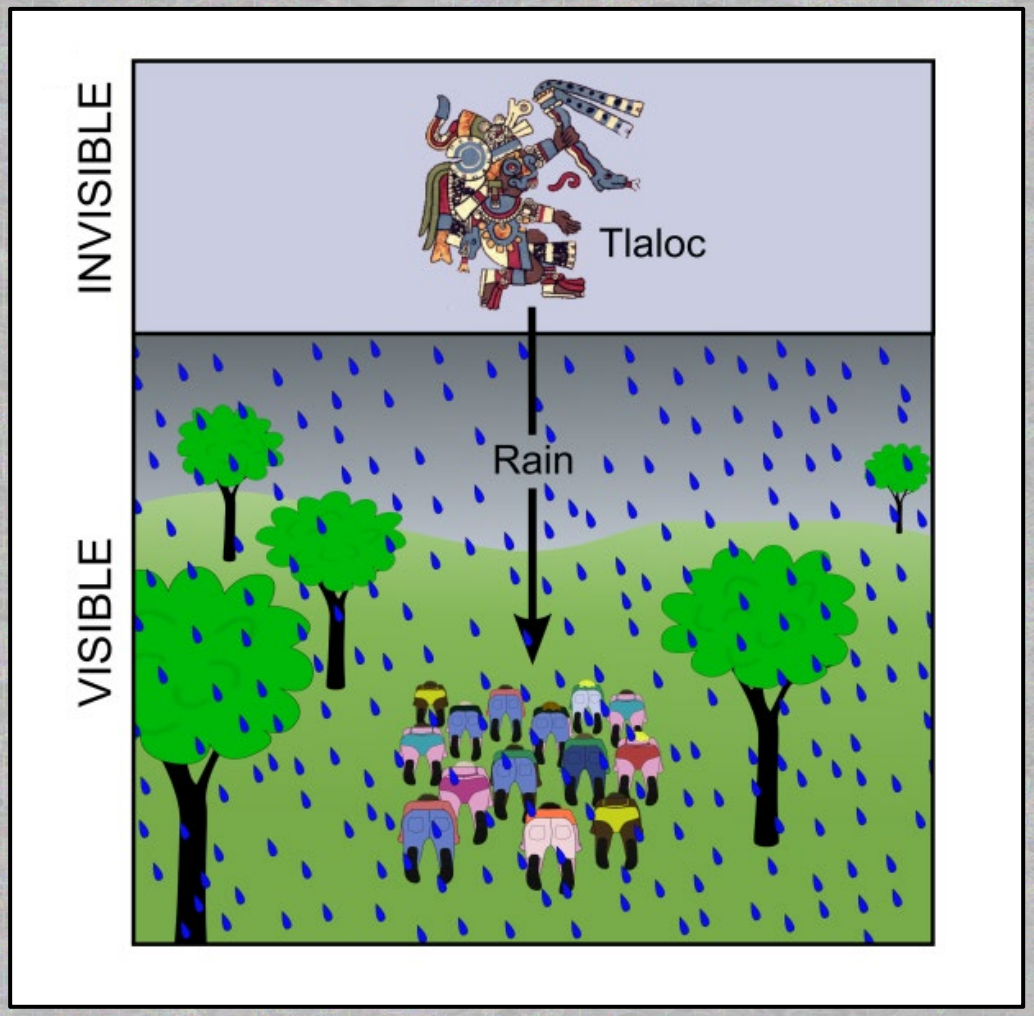
Invisible Explanations

- Examples of invisible explanations:
  - God(s).
  - X-rays.
  - Atoms.
  - Modern physics.

Tlaloc



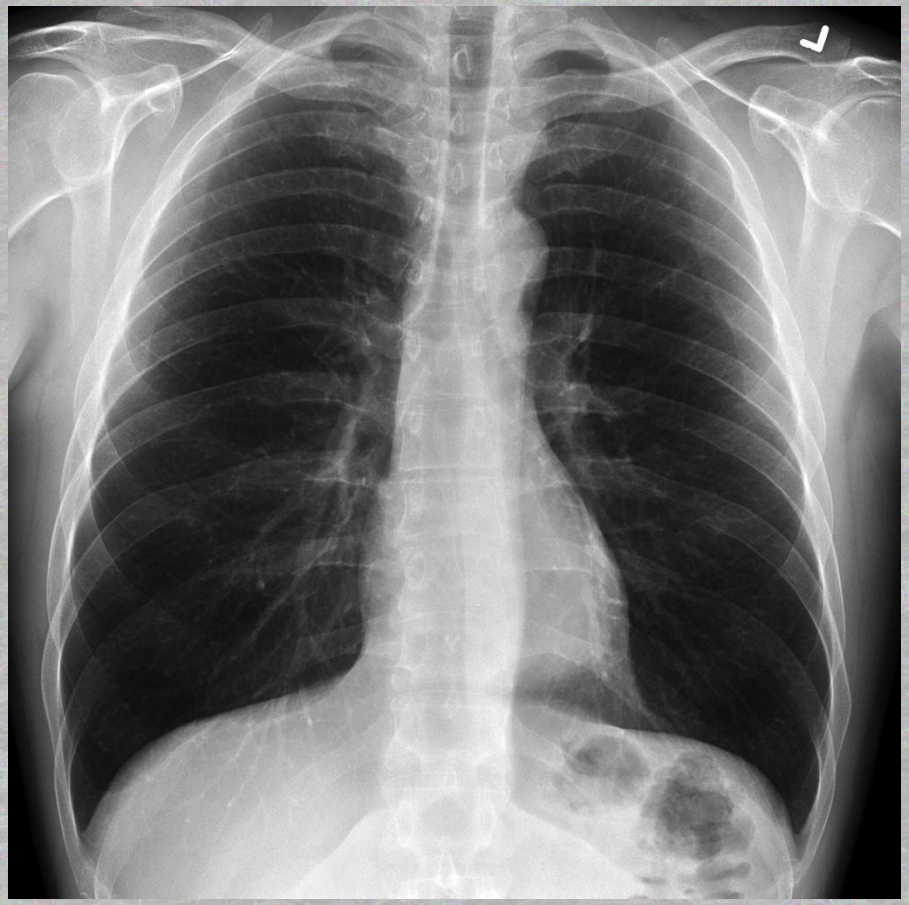
Tlaloc



X-rays

- X-rays are invisible waves that were posited to explain the appearance of patterns on photographic plates.
- These patterns can easily be explained if there is a form of radiation that cannot be perceived with the human eye.
- We only perceive the *effects* of X-rays, never the X-rays themselves.
- X-rays are invisible explanations.

X-rays



Atoms

- In the 17th Century there was a renaissance of atomism.
- Atoms in the void were used to explain the world that we see around us – for example Boyle's law.
- 17th Century scientists could not see the atoms.
- The atoms were invisible explanations.

Modern Physics

- Modern physics posits many entities to explain phenomena that we see around us:
  - Four-dimensional spacetime.
  - Ten-dimensional superstrings.
  - Wave-particles.
  - Etc.
- We have no direct experience of any of them.
- They are invisible explanations.

The Physical World

- Physical entities are invisible explanations for the world that we experience around us.
- We can measure the physical world.
- We have never seen the physical world.
- The predictive success of science convinces us that the physical world is really out there.

The Invisible Physical World

Modern physics, therefore, reduces matter to a set of events which proceed outward from a centre. If there is something further in the centre itself, we cannot know about it, and it is irrelevant to physics. … Physics is mathematical, not because we know so much about the physical world, but because we know so little: it is only its mathematical properties that we can discover. For the rest, our knowledge is negative. In places where there are no eyes or ears or brains there are no colours or sounds, but there are events having certain characteristics which lead them to cause colours and sounds in places where there are eyes, ears and brains. We cannot find out what the world looks like from a place where there is nobody, because if we go to look there will be somebody there; the attempt is as hopeless as trying to jump on one's own shadow.

Bertrand Russell, *An Outline of Philosophy*, p. 163

Negative Definition of Consciousness

- Physical world is invisible and can be described mathematically.
- Consciousness is everything in our naïve encounters with the world that is non-physical.
- Colours, smells sounds, etc. are non-physical.
- These are properties of consciousness.

Naïve Realism

Invisible Physical World

Consciousness

Bubble of Experience

- Our experiences are not like photographs.
- We look out from our bodies into a world.
- Our experiences change every ~200 ms.
- I have described this using the idea of a *bubble of experience*.
- This is a bubble of space, roughly centred on our bodies, that contains colours, smells, sounds, body sensations etc.

Bubble of Experience

Positive Definition of Consciousness

- Consciousness is a bubble of experience.
- Phenomenologists, such as Husserl and Merleau-Ponty describe the structures of our bubbles of experience.
- The invisible physical world is used to explain and predict the regularities in our bubbles of experience.

Development of the Modern Concept of Consciousness

- The modern concept of consciousness emerged ~300 years ago.
- Atoms lacked properties that we encountered in naïve realism - colour, smell, taste, etc.
- To accommodate this, Locke and Galileo distinguished two types of properties:
  - **Primary qualities** - movement, size, shape etc.
  - **Secondary qualities** - colour, smell, taste etc.

Development of the Modern Concept of Consciousness

- **Primary qualities** are properties of atoms.
- **Secondary qualities** are properties of a second substance called consciousness.
- The modern concept of consciousness was invented to contain the properties that were removed from the physical world by modern science.

Development of the Modern Concept of Consciousness

Now I say that whenever I conceive any material or corporeal substance, I immediately feel the need to think of it as bounded, and as having this or that shape; as being large or small in relation to other things, and in some specific place at any given time; as being in motion or at rest; as touching or not touching some other body; and as being one in number, or few, or many. From these conditions I cannot separate such a substance by any stretch of my imagination. But that it must be white or red, bitter or sweet, noisy or silent, and of sweet or foul odour, my mind does not feel compelled to bring in as necessary accompaniments. Without the senses as our guides, reason or imagination unaided would probably never arrive at qualities like these. Hence I think that tastes, odours, colors, and so on are no more than mere names so far as the object in which we place them is concerned, and that they reside only in the consciousness. Hence if the living creature were removed, all these qualities would be wiped away and annihilated. But since we have imposed on them special names, distinct from those of the other and real qualities mentioned previously, we wish to believe that they really exist as different from those.

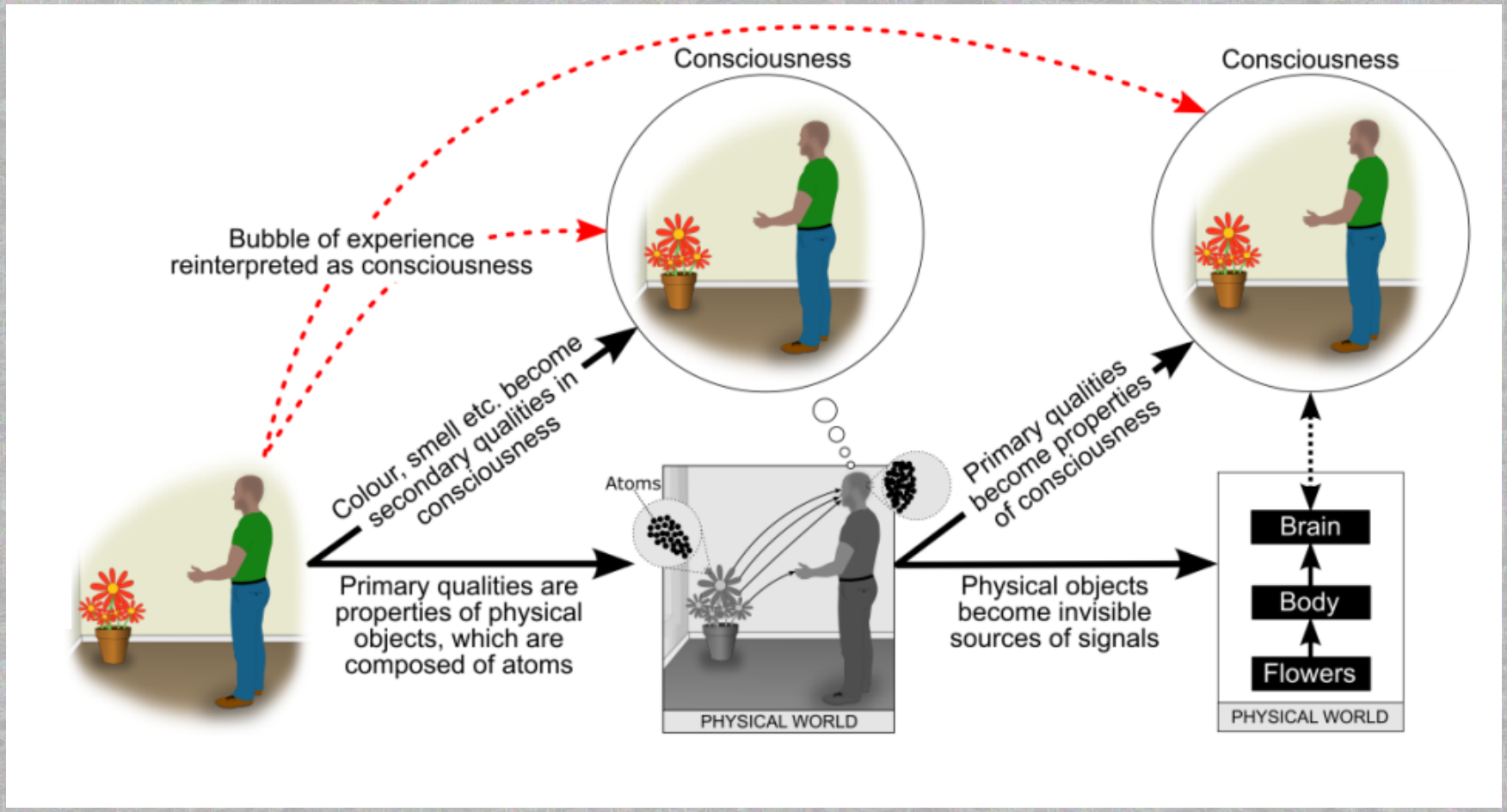
Gallileo Galilei, *Discoveries and Opinions of Galileo*, p. 274

Development of the Modern Concept of Consciousness

Now I say that whenever I conceive any material or corporeal substance, I immediately feel the need to think of it as bounded, and as having this or that shape; as being large or small in relation to other things, and in some specific place at any given time; as being in motion or at rest; as touching or not touching some other body; and as being one in number, or few, or many. From these conditions I cannot separate such a substance by any stretch of my imagination. But that it must be white or red, bitter or sweet, noisy or silent, and of sweet or foul odour, my mind does not feel compelled to bring in as necessary accompaniments. Without the senses as our guides, reason or imagination unaided would probably never arrive at qualities like these. Hence I think that **tastes, odours, colors, and so on are no more than mere names so far as the object in which we place them is concerned, and that they reside only in the consciousness**. Hence if the living creature were removed, all these qualities would be wiped away and annihilated. But since we have imposed on them special names, distinct from those of the other and real qualities mentioned previously, we wish to believe that they really exist as different from those.

Gallileo Galilei, *Discoveries and Opinions of Galileo*, p. 274

Development of the Modern Concept of Consciousness



Linguistic Evidence

Two intriguing facts. First, the terms ‘mind’ and ‘conscious(ness)’ are notoriously difficult to translate into some other languages. Second, in English (and other European languages) one of these terms – ‘conscious’ and its cognates – is in its present range of senses scarcely three centuries old. … In ancient Greek there is nothing corresponding to either ‘mind’ or ‘consciousness’ … In Chinese, there are considerable problems in capturing ‘conscious(ness)’.

Kathleen Wilkes, ‘___, yìshì, duh, um, and consciousness’, pp. 16-7

Summary

- There is a close connection between modern science and the modern concept of consciousness.
- These concepts have co-evolved over the last 300 years.
- Consciousness is everything that we experience as we interact with the world (everything in naïve realism).
- The physical world is an invisible explanation for regularities in consciousness.

Thought Experiments and Imagination

Hard Problem of Consciousness

- People (particularly philosophers) often try to use thought experiments and imagination to identify the relationship between consciousness and the physical world.
- Often end up with the so-called ‘hard problem of consciousness’.

Hard Problem of Consciousness

How is it possible for conscious states to depend upon brain states? How can technicolour phenomenology arise from soggy grey matter? What makes the bodily organ we call the brain so radically different from other bodily organs, say the kidneys - the body parts without a trace of consciousness? How could the aggregation of millions of individually insentient neurons generate subjective awareness? We know that brains are the de facto causal basis of consciousness, but we have, it seems, no understanding whatever of how this can be so. It strikes us as miraculous, eerie, even faintly comic. Somehow, we feel, the water of the physical brain is turned into the wine of consciousness, but we draw a total blank on the nature of this conversion. Neural transmissions just seem like the wrong kind of materials with which to bring consciousness into the world, but it appears that in some way they perform this mysterious feat.

Colin McGinn, Can We Solve the Mind-Body Problem?

Hard Problem of Consciousness

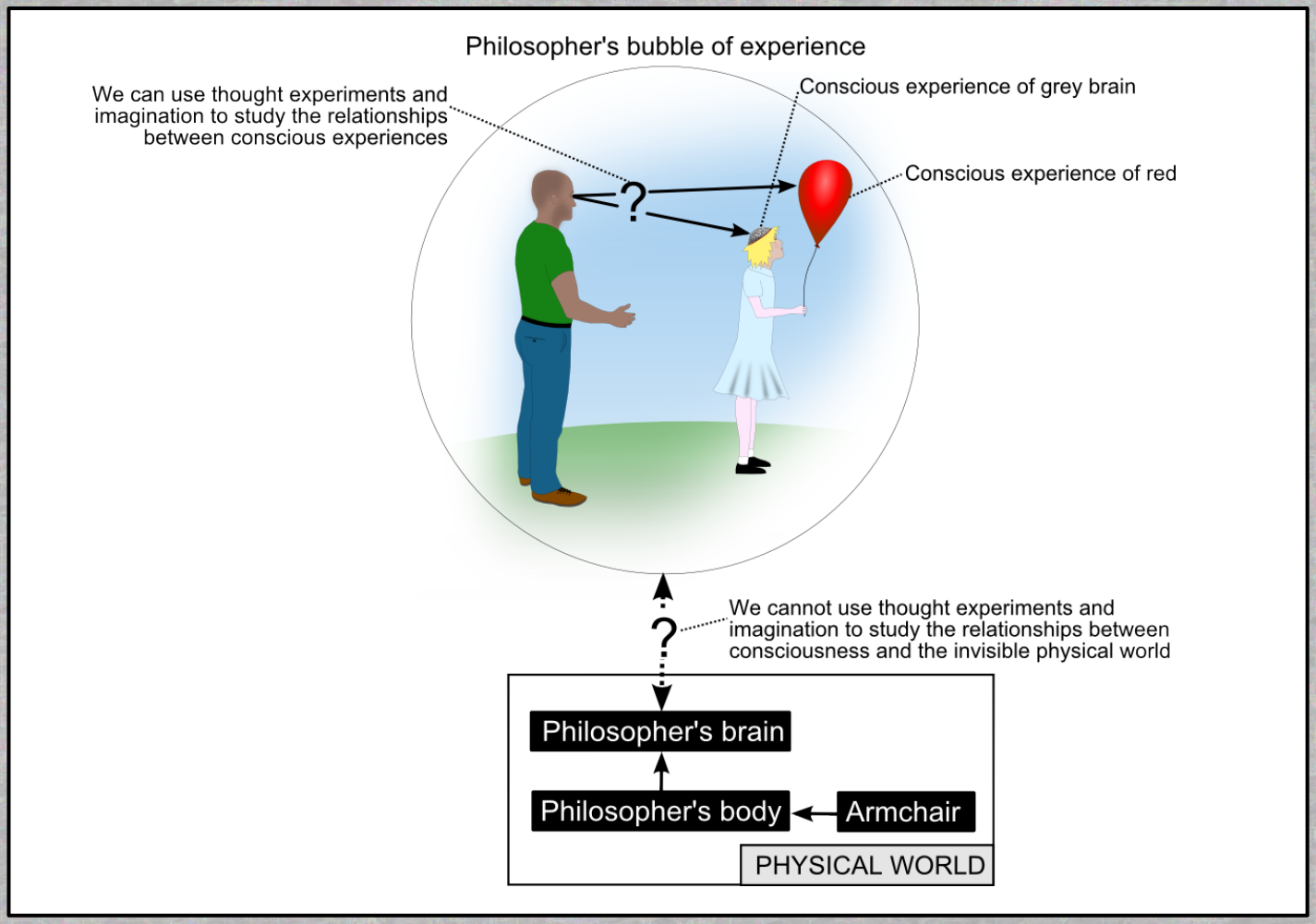
How is it possible for conscious states to depend upon brain states? **How can technicolour phenomenology arise from soggy grey matter?** What makes the bodily organ we call the brain so radically different from other bodily organs, say the kidneys - the body parts without a trace of consciousness? How could the aggregation of millions of individually insentient neurons generate subjective awareness? We know that brains are the de facto causal basis of consciousness, but we have, it seems, no understanding whatever of how this can be so. It strikes us as miraculous, eerie, even faintly comic. Somehow, we feel, the water of the physical brain is turned into the wine of consciousness, but we draw a total blank on the nature of this conversion. Neural transmissions just seem like the wrong kind of materials with which to bring consciousness into the world, but it appears that in some way they perform this mysterious feat.

Colin McGinn, Can We Solve the Mind-Body Problem?

Hard Problem of Consciousness

- The ‘hard problem of consciousness’ becomes a pseudo problem when you realise that the physical world is invisible.
- The relationships between conscious experiences cannot help us to understand the relationship between conscious experiences and the invisible physical world.

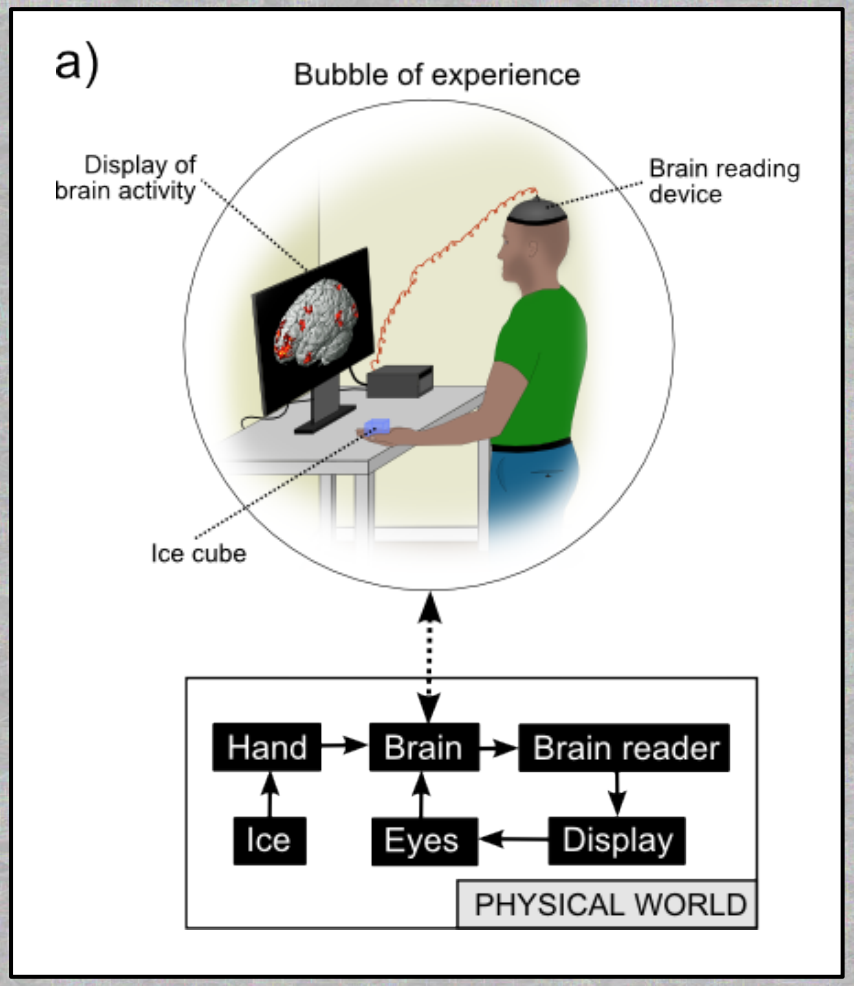
Hard Problem of Consciousness



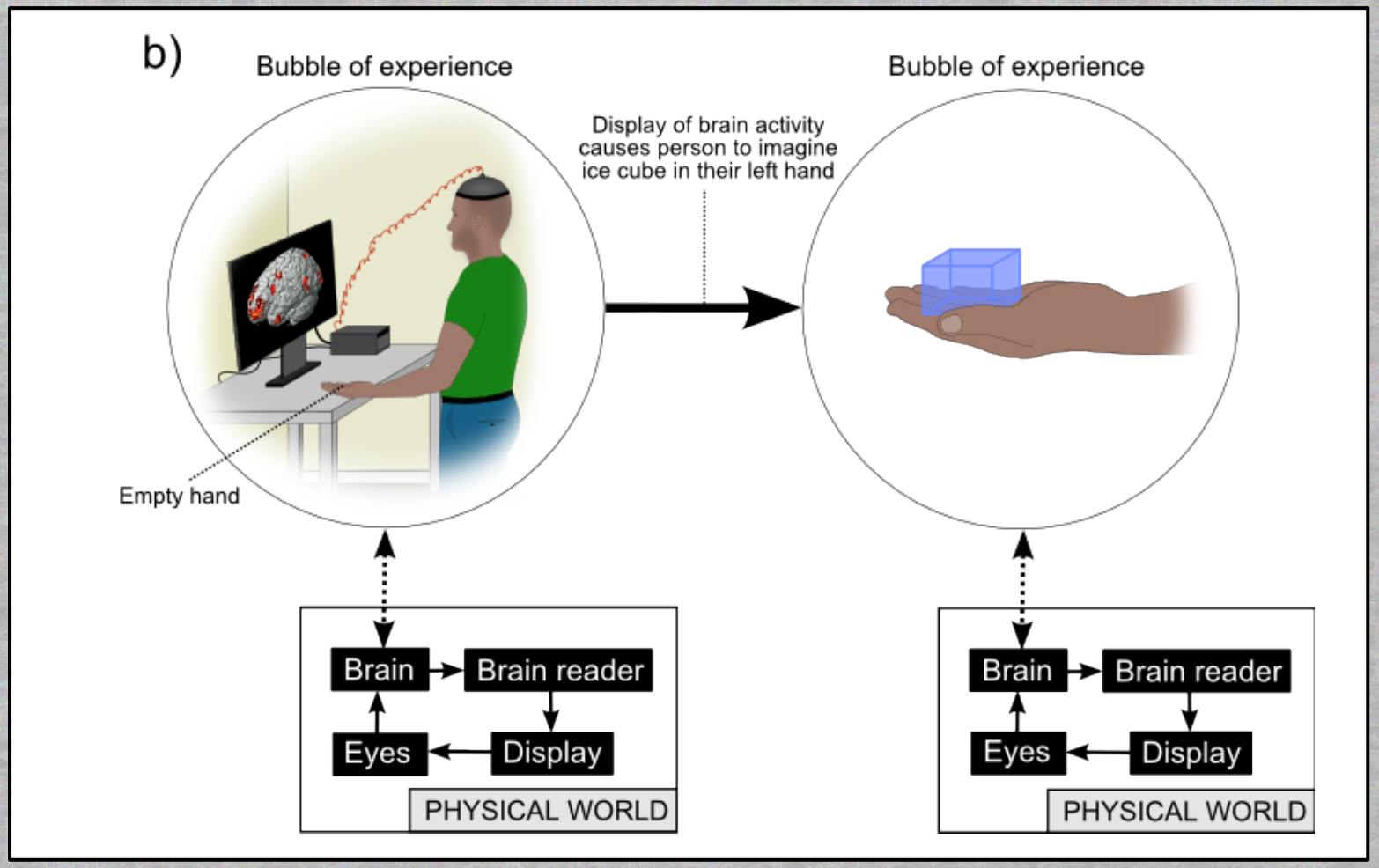
Regularities in Conscious Experience

- We can use regularities in conscious experiences to make inferences about regularities in the physical world.
- In theory we could learn the relationship between consciously experience brain activity and other conscious experiences.
- Have not been exposed to enough data to do this yet.

Regularities in Conscious Experience



Regularities in Conscious Experience



Summary

- The ‘hard problem of consciousness’ is a mistaken attempt to use the relationships between conscious experiences to understand the relationship between conscious experiences and the invisible physical world.
- We can make inferences from conscious experiences of brains to other conscious experiences.
- This is difficult because we have not been exposed to this relationship and we have a limited ability to perceive and learn complex patterns.

SCIENCE OF CONSCIOUSNESS

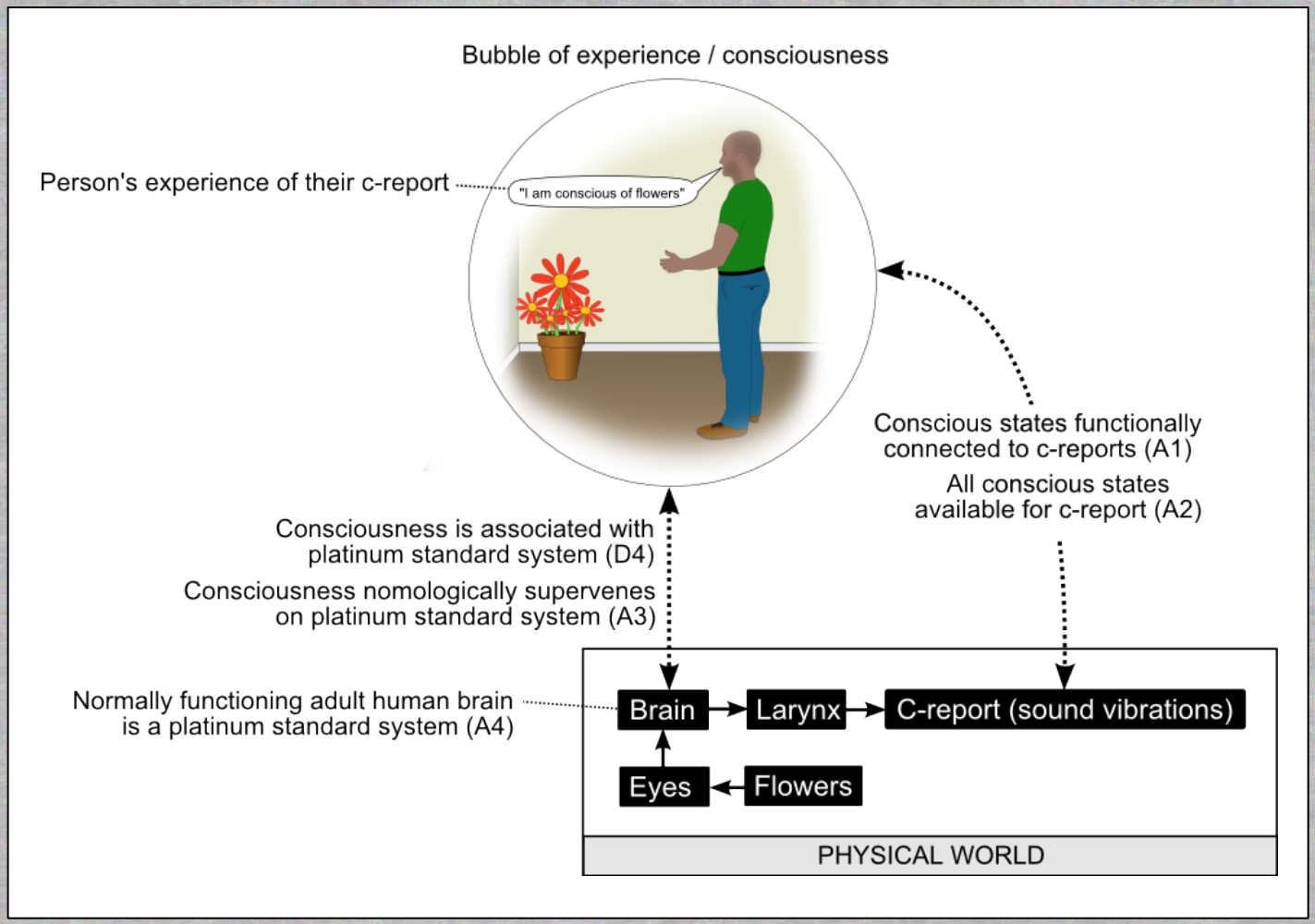
Science of Consciousness

- We can develop a scientific understanding of the relationship between consciousness and the physical world.
- Applications of the science of consciousness:
  - Diagnosis of coma patients.
  - Repair of damaged consciousness.
  - Informed choices about animal welfare.
  - Human-AI communication.
  - Robotics.
  - Conscious machines.
  - Eternal life (uploading into cloud)

Measurement of Consciousness

- To study consciousness scientifically we need to measure it.
- Consciousness is measured through first-person reports.
- This raises a number of philosophical problems.
- These can be handled with *assumptions*.
- The science of consciousness is considered to be true *given these assumptions*.

Assumptions for the Measurement of Consciousness



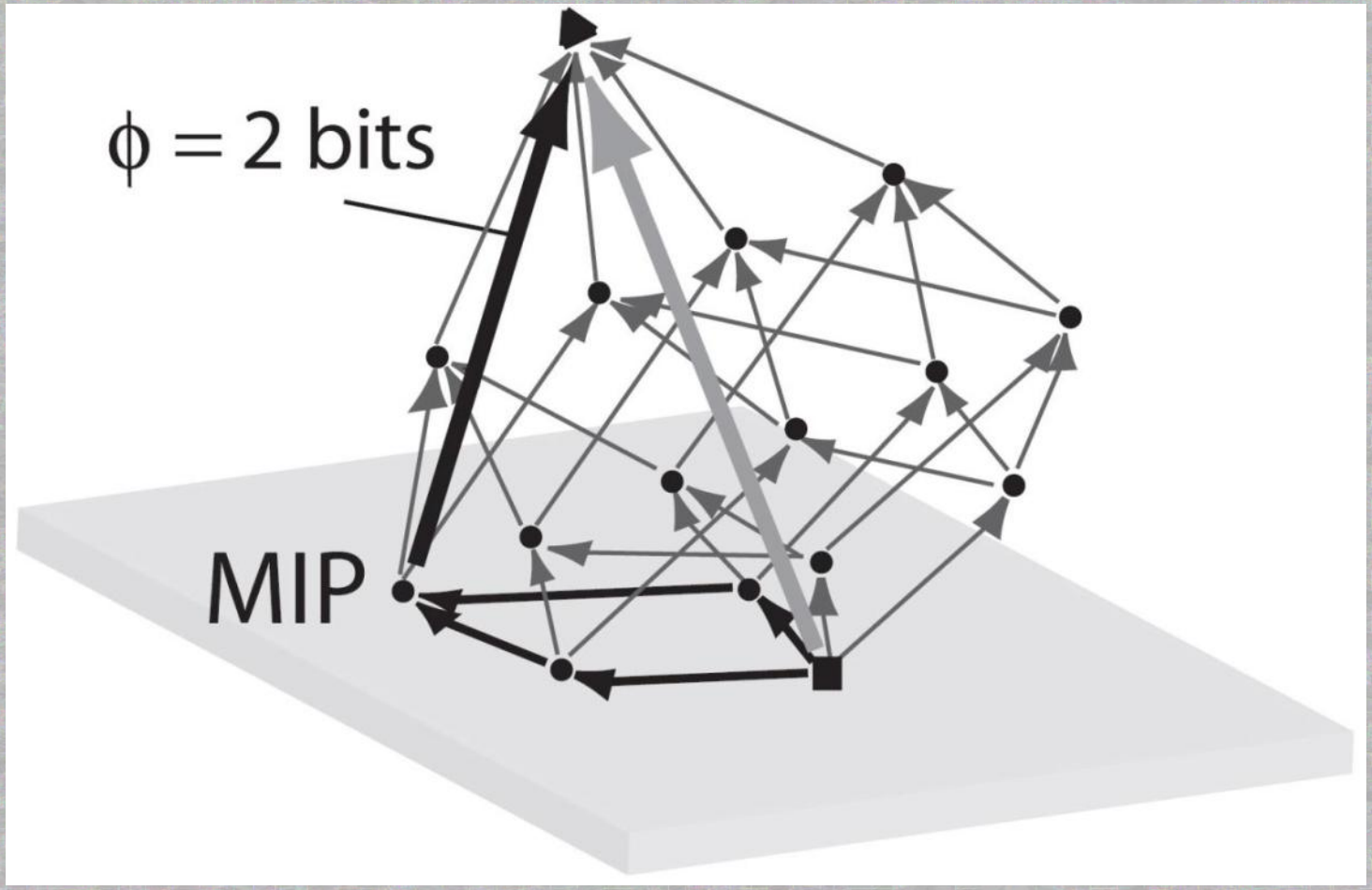
Description of Consciousness

- Consciousness cannot be described in natural language, which is:
  - Context-bound
  - Ambiguous
  - Not applicable to infants, bats, robots, etc.
  - Not mathematically tractable.

C-description

- Need a precise formal way of describing consciousness that is applicable to any system.
- Will refer to this as a *c-description*.
- Possible methods include:
  - XML/LMNL
  - High dimensional qualia (Balduzzi and Tononi)
  - Category theory

IIT: C-description of Conscious State

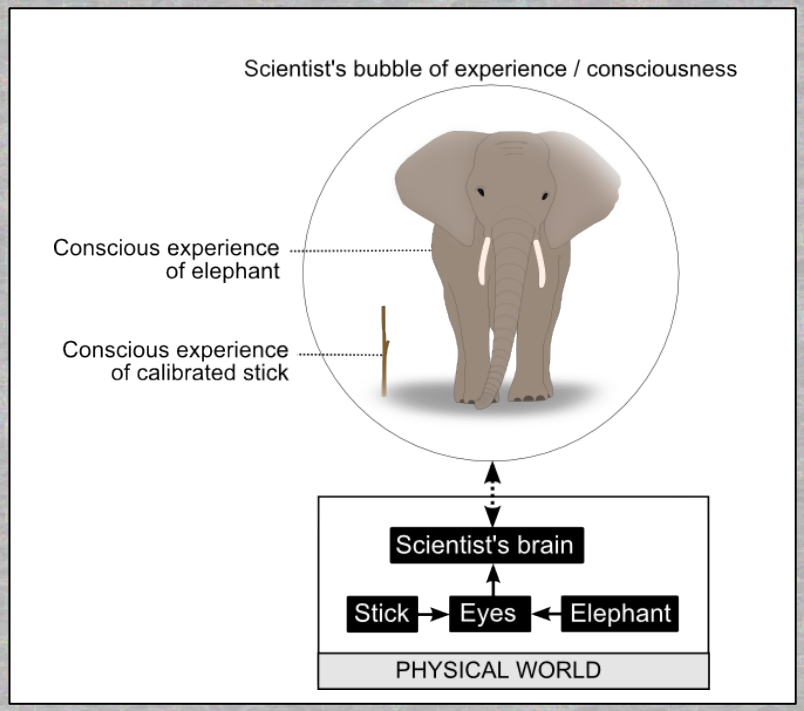


XML C-description of Conscious State

```
- <mental_state>
  - <omc_scale>
      <rating>0.427</rating>
      <version>0.6</version>
    </omc_scale>
  - <physical_description>
    - <firing_neuron>
        <id>120621</id>
      </firing_neuron>
    </physical_description>
  - <cluster>
      <id>207193</id>
      <type>phi</type>
      <amount>75.1173</amount>
    </cluster>
  - <representations>
    - <output>
      - <neuron>
          <id>127936</id>
        </neuron>
        <mutual_information>0.993765</mutual_information>
        <human_description>Proprioception / motor output</human_description>
        <physical_description>N/A</physical_description>
      </output>
    - <input>
      - <neuron>
          <id>104137</id>
        </neuron>
        <mutual_information>0.99159</mutual_information>
        <human_description>Red / blue visual input</human_description>
        <physical_description>700/450 nm electromagnetic waves</physical_description>
      </input>
    </representations>
  - <phenomenal_predictions>
      <tononi>0</tononi>
      <aleksander>0.99159</aleksander>
      <metzinger>75.1173</metzinger>
    </phenomenal_predictions>
  </mental_state>
- <mental_state>
```

Measurement of the Physical World

- The scientist has a conscious experience in which an object interacts with a calibrated object.
- He/she observes the result and extracts a number.



Description of the Physical World

- The number that is extracted through a measurement procedure is attributed to an object in the physical world.
- 3 metres is the *height of an elephant.*
- Objects are tightly defined in physics and chemistry.
- They are not tightly defined in biology.

P-description

- We want a science of consciousness that can make predictions about the consciousness of arbitrary systems (bats, robots, rocks etc.)
- A science of consciousness based on biological neurons will not be able to say anything about the consciousness of systems based on synthetic neurons.
- Need a precise formal description of the spatiotemporal physical structures that are linked to consciousness.
- Will be referred to as a *p-description*.

Physical, Informational or Computational States?

- Researchers on consciousness have focused on three different features of the physical world that might be linked to consciousness.
  - Physical states.
  - Informational states.
  - Computational/functional states.
- Only physical states are objective.
- Computations, functions and information are observer relative and cannot be part of a science of consciousness.

Neural Correlates of Consciousness

- Has been a lot of scientific work on the neural correlates of consciousness.
- Look for synchronization, connection patterns, etc. that are present when consciousness is present and absent when consciousness is absent.

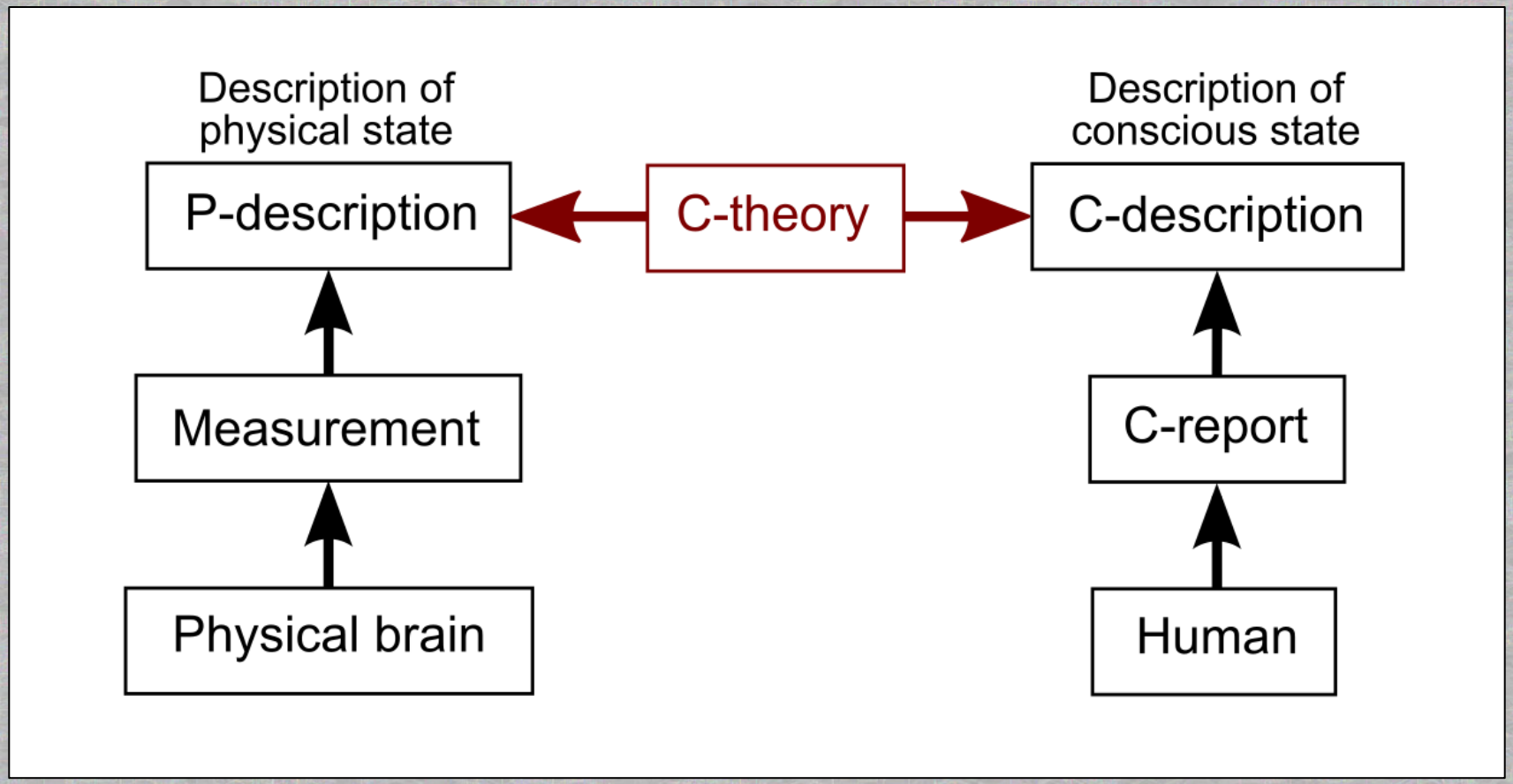
From Correlates to Theories of Consciousness

- Research on the neural correlates of consciousness yields useful data.
- Our final theory of consciousness will not be a long list of correlations between consciousness and the physical world.
- We want a compact mathematical description of the relationship between consciousness and the physical world

Theory of Consciousness (C-theory)

- A c-theory is a mathematical description of the relationship between measurements of consciousness (c-descriptions) and measurements of the physical world (p-descriptions).
- It can generate c-descriptions from p-descriptions.
- It can generate p-descriptions from c-descriptions.

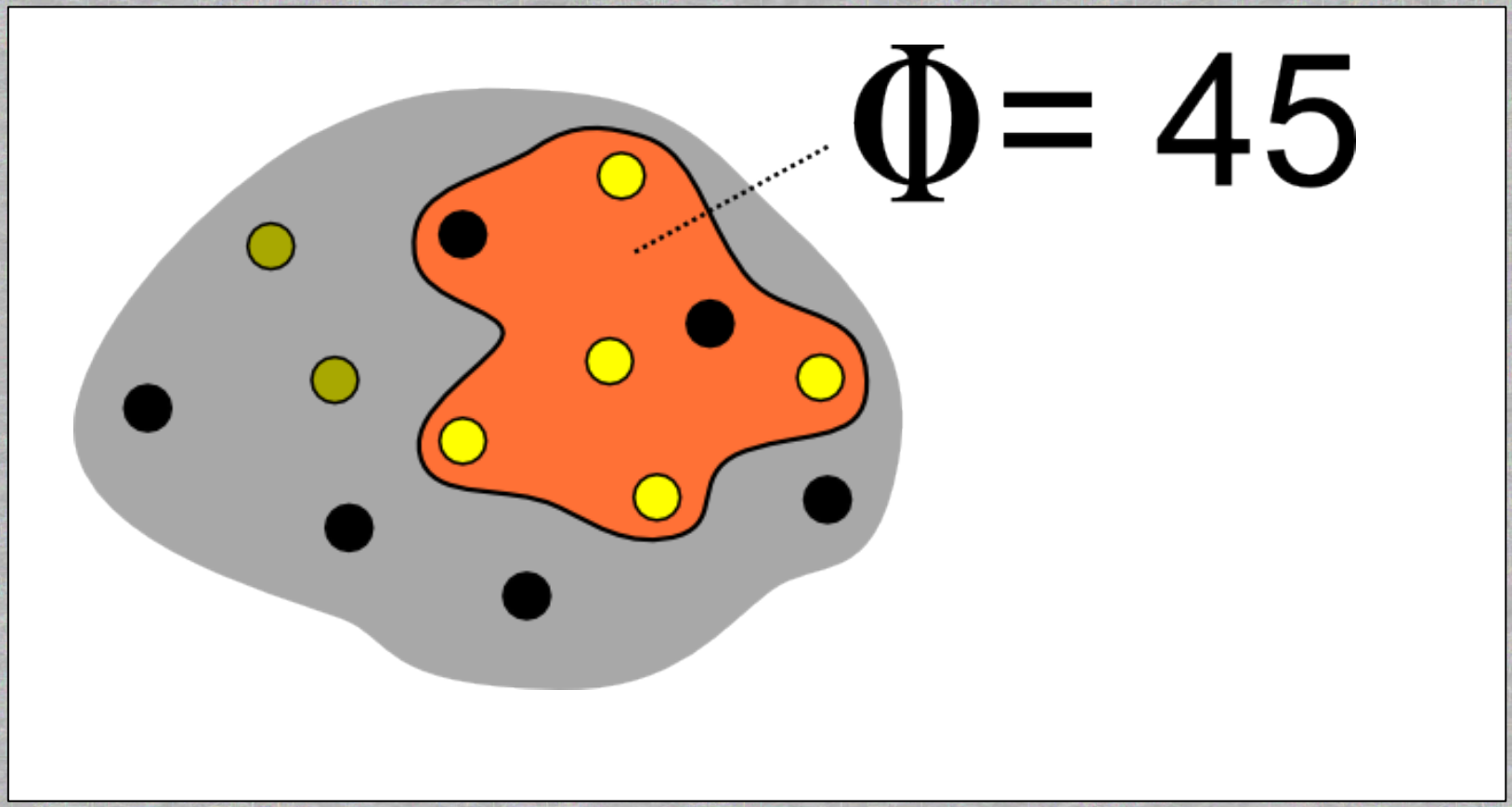
Mathematical C-theory



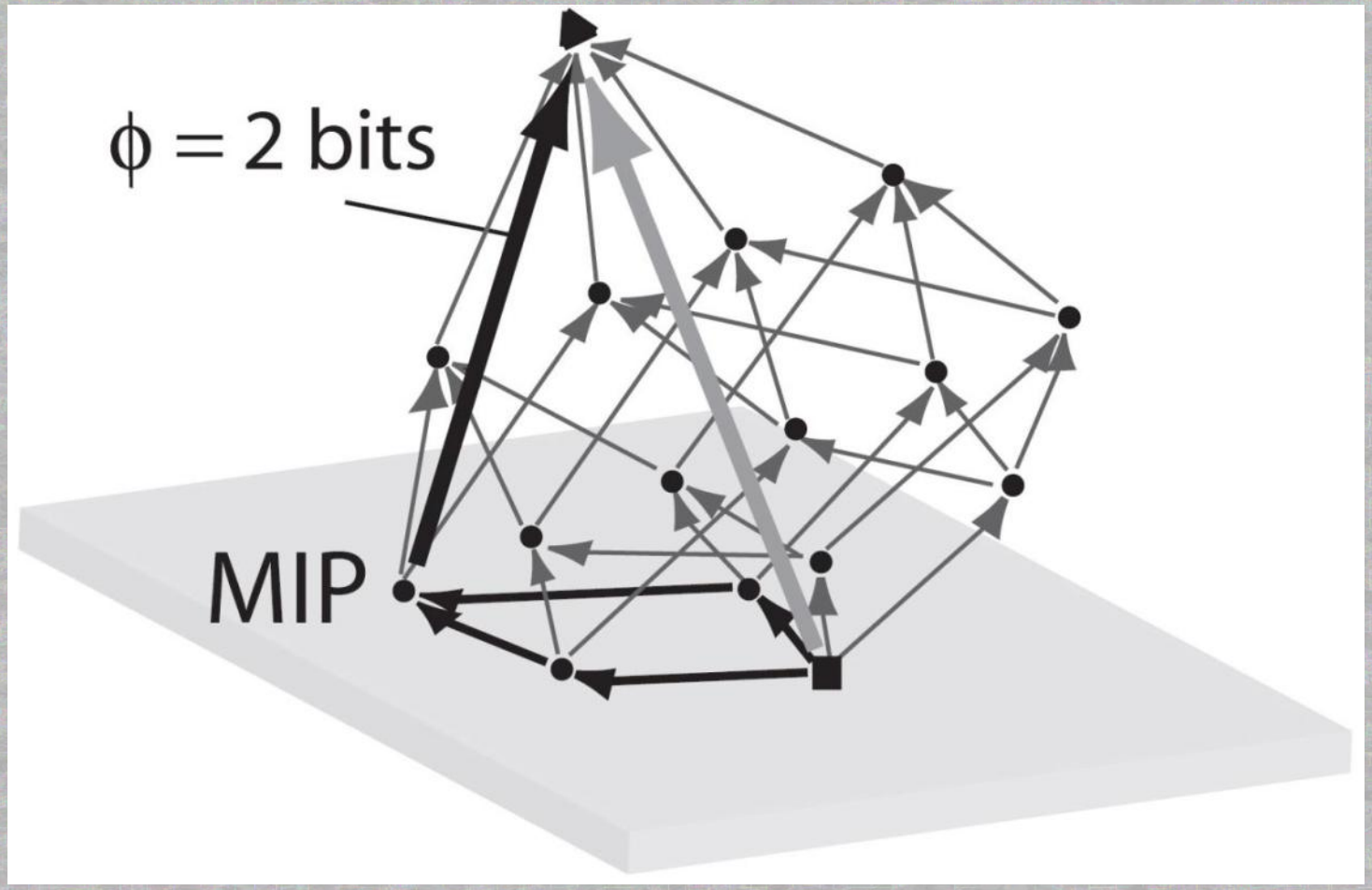
Example: Tononi's Information Integration Theory of Consciousness (IIT)

- Tononi's IIT is the closest thing to a c-theory that we have so far.
- A mathematical algorithm links a description of the physical world to a description of consciousness.
- A conscious state (a quale) is c-described using a high dimensional mathematical structure.

IIT: The Conscious Part of the System



IIT: Description of the Contents of Consciousness



Limitations of IIT

- IIT is very popular right now.
- It has the correct ***form*** of a scientific theory of consciousness.
- However, it has serious limitations:
  - Based on subjective information.
  - Severe performance limitations.
  - No compelling evidence.

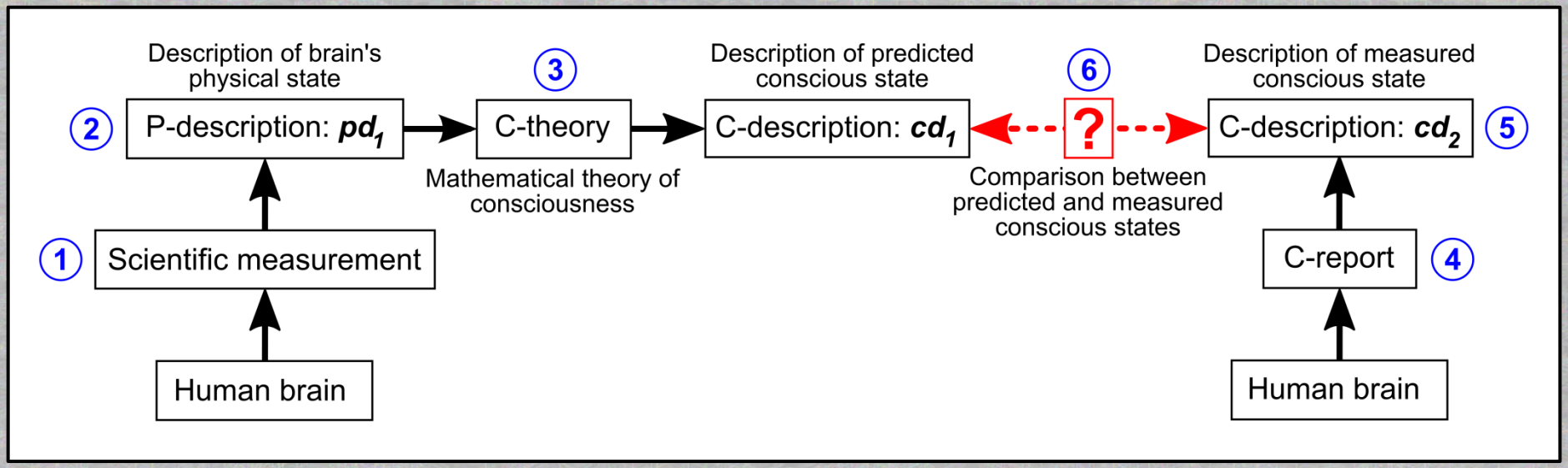
Discovery of Scientific Theories of Consciousness

- Traditionally *people* have identified regularities in the physical world (Newton, Einstein, etc.).
- We generally assume that physical regularities are simple enough to be found by humans.
- This is likely to be the wrong approach for the discovery of c-theories.
- Will probably have to use AI/machine learning to discover mathematical relationships between consciousness and the physical world.

Predictions about Consciousness

- Mathematical theories of consciousness can generate predictions about consciousness.
- These predictions can be used to test the theories.

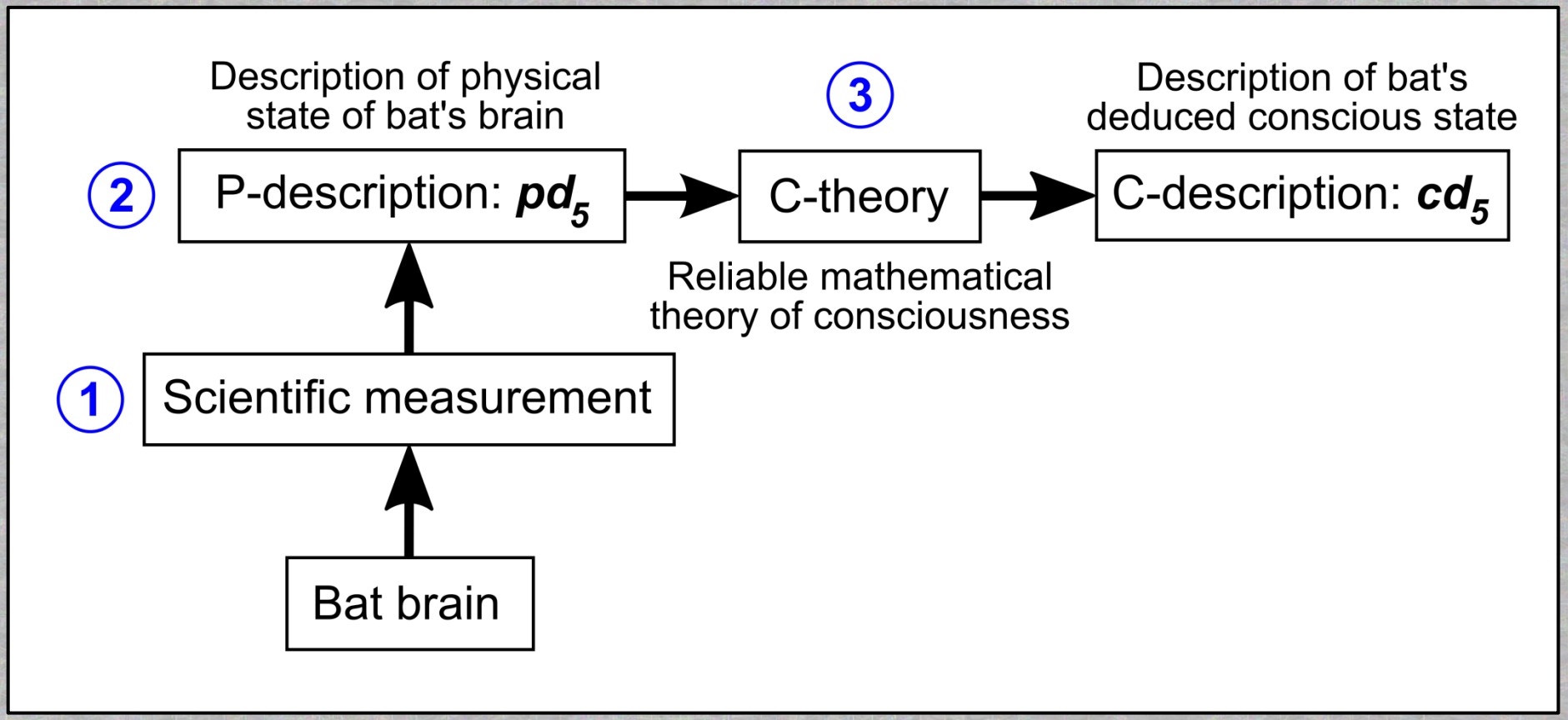
Prediction about Consciousness



Deductions about Consciousness

- We make **deductions** about the consciousness of a system when consciousness cannot be measured through first-person reports.
- For example:
  - Infants.
  - Animals.
  - Robots.

Deduction of the Consciousness of a Bat



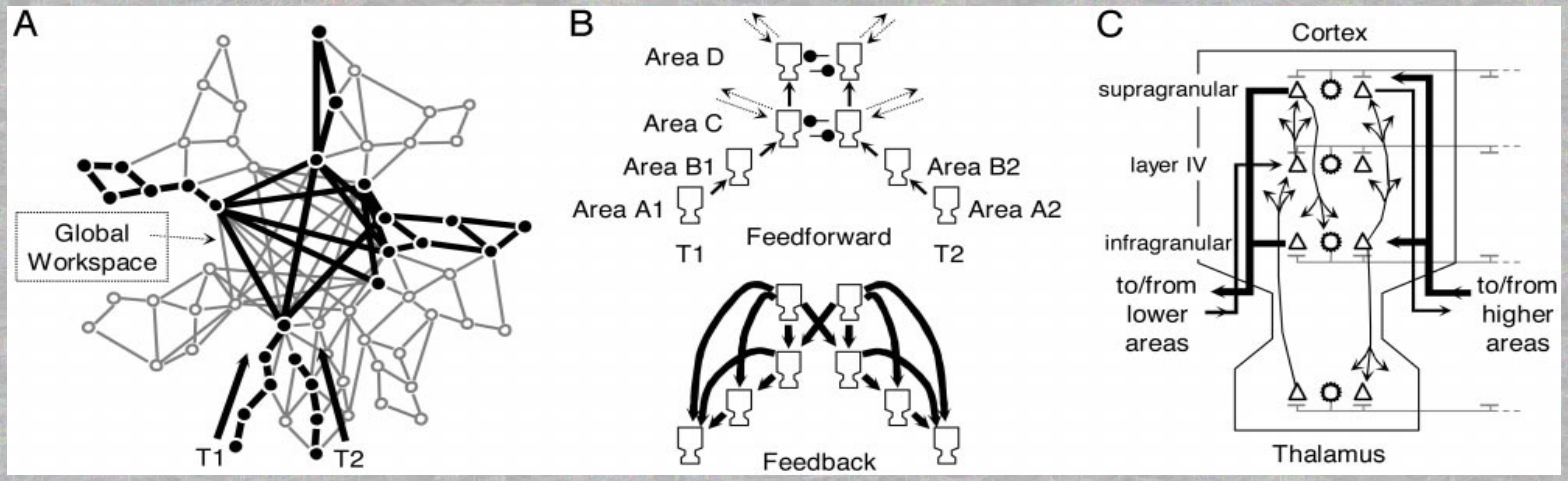
Summary

- Science of consciousness:
  - Measure consciousness.
  - Measure physical world.
  - Use machine learning to discover mathematical relationships between the two sets of measurements.
- Results of the science of consciousness depend on philosophical assumptions.
- Information, computations and functions are subjective - focus on physical properties of the world.

Models of Consciousness

Models of the Correlates of Consciousness in Neuroscience

- Neuroscientists build models to help them to understand potential neural correlates of consciousness.



Machine Consciousness

- Models of the correlates of consciousness and models of consciousness are used to build intelligent machines that are potentially conscious.
- Complex field with several overlapping objectives.

Types of Machine Consciousness

- **MC1**. Machines with the same external behaviour as conscious humans.
- **MC2**. Computer models of the correlates of consciousness.
- **MC3**. Computer models of consciousness.
- **MC4**. Machines that really have conscious experiences.

Types of Machine Consciousness

- **MC1**. Machines with the same external behaviour as conscious humans.
- **MC2**. Computer models of the correlates of consciousness.
- **MC3**. Computer models of consciousness.
- **MC4**. Machines that really have conscious experiences.

Conscious Human Behaviours

- Humans have characteristic behaviours when they are conscious.
- For example:
  - Alertness.
  - Response to novel situations.
  - Inward execution of sequences of problem-solving steps.
  - Learning.
  - Response to verbal commands.
  - Delayed response to stimuli.

MC1 Machine Consciousness

- A machine is MC1 conscious if it is producing similar external behaviour to a conscious human.
- Many artificially intelligent machines are already MC1 conscious to some extent.
- For example, humans can only play Atari video games, Go or Jeopardy! when they are conscious.
- MC1 machine consciousness is part of artificial general intelligence (AGI).

IBM Watson


PIENSE
THINK
ΣΚΕΨΟΥ
DENKE
PENSER
$300,000
$1,000,000
$200,000
KEN
WATSON
BRAD

Types of Machine Consciousness

- **MC1**. Machines with the same external behaviour as conscious humans.
- **MC2**. Computer models of the correlates of consciousness.
- **MC3**. Computer models of consciousness.
- **MC4**. Machines that really have conscious experiences.

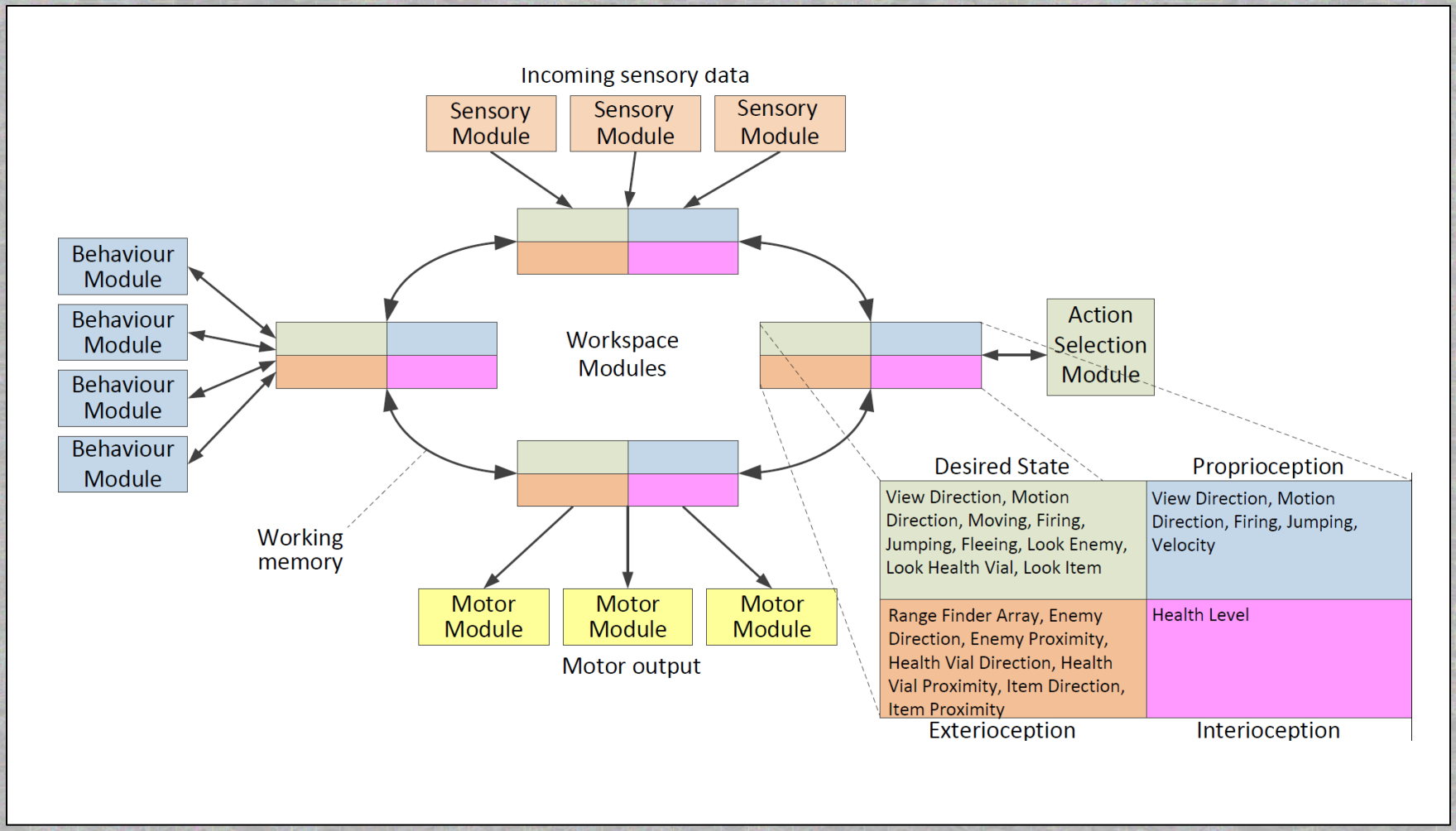
Models of the Correlates of Consciousness

- MC2 machine consciousness is the construction of:
  - Models of the neural correlates of consciousness.
  - Models of the cognitive correlates of consciousness.

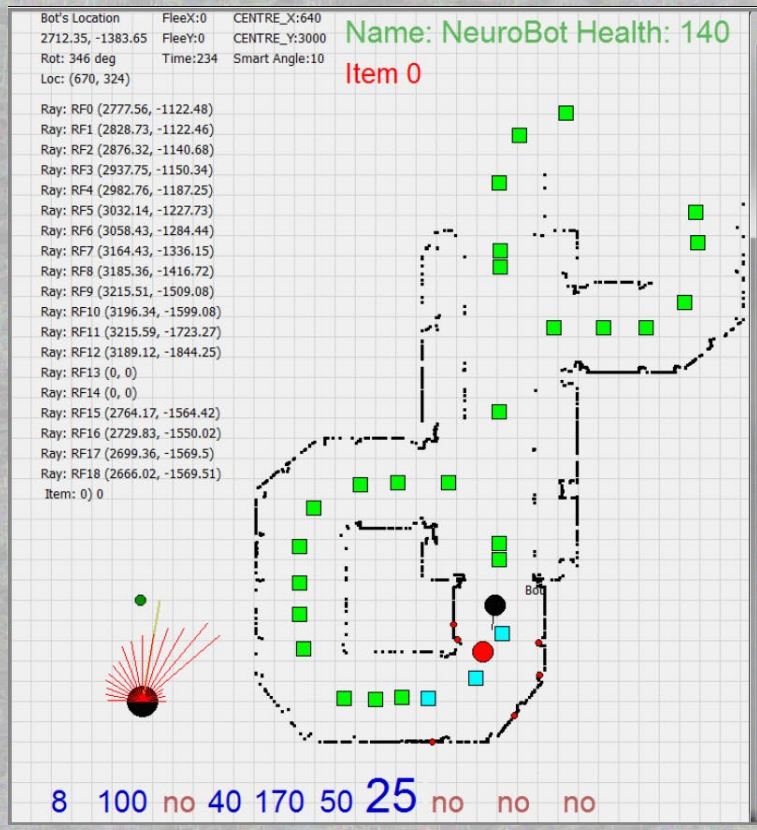
NeuroBot

- Neural implementation of global workspace.
- Controlled an avatar in the Unreal Tournament 2004 game environment.
- 20,000 neurons; 1.5 million connections.
- Implemented by Zafeirios Fountas.

Network Architecture



Unreal Tournament 2004



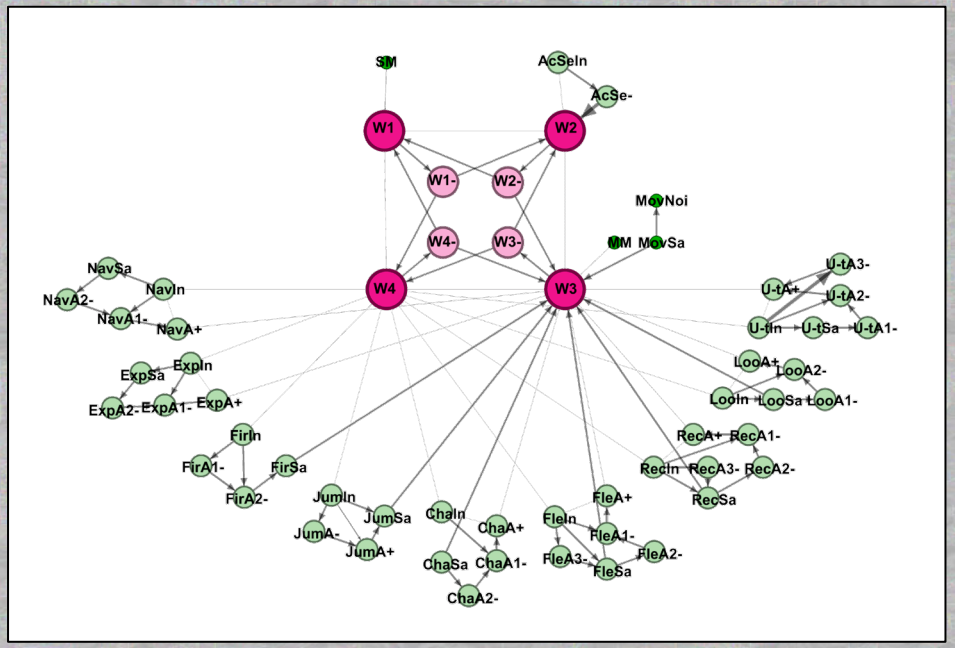
Results

- Close second in 2011 Botprize competition.
- Developed a metric ᛗ (*Mannaz*) to measure the humanness of behaviour and carried out a series of experiments comparing humans with each other and with a couple of bots.

Statistical Measure	h1	h2	h3	h4	h5	NeuroBot	Hunter
Exploration Factor	80	80	80	20	80	60	20
Stationary Time	60	40	60	20	40	60	0
Path Entropy	80	80	80	20	80	0	0
Average Health	80	80	80	20	80	20	20
Number of Kills	60	60	80	20	80	60	0
Number of Deaths	40	80	60	20	80	0	0
ᛗ	67	70	73	20	73	33	7

Analysis of NeuroBot's Network

- Analyzed network's structural connectivity using different graph theory measures to see if its information-processing was similar to a global workspace.



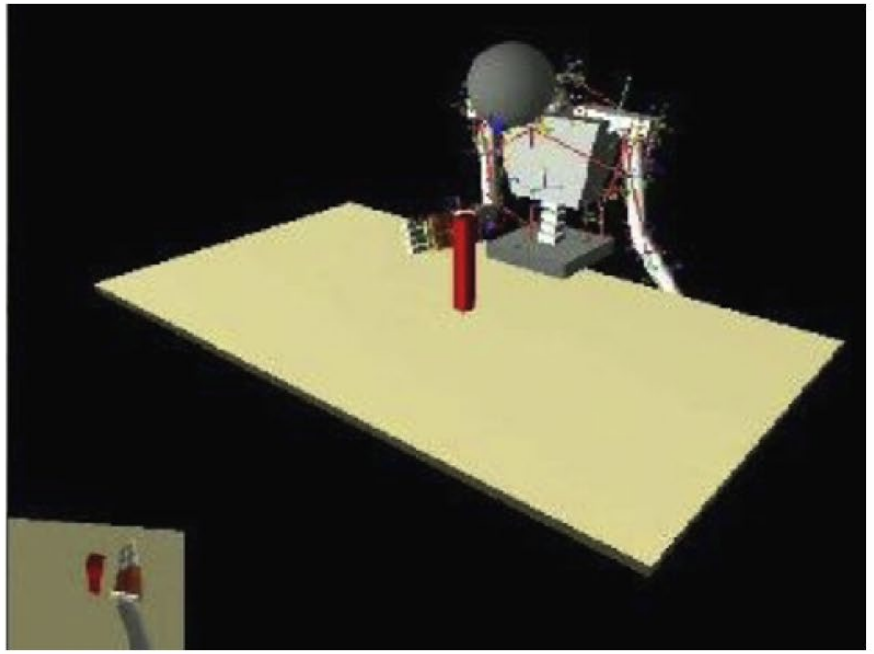
Types of Machine Consciousness

- **MC1**. Machines with the same external behaviour as conscious humans.
- **MC2**. Computer models of the correlates of consciousness.
- **MC3**. Computer models of consciousness.
- **MC4**. Machines that really have conscious experiences.

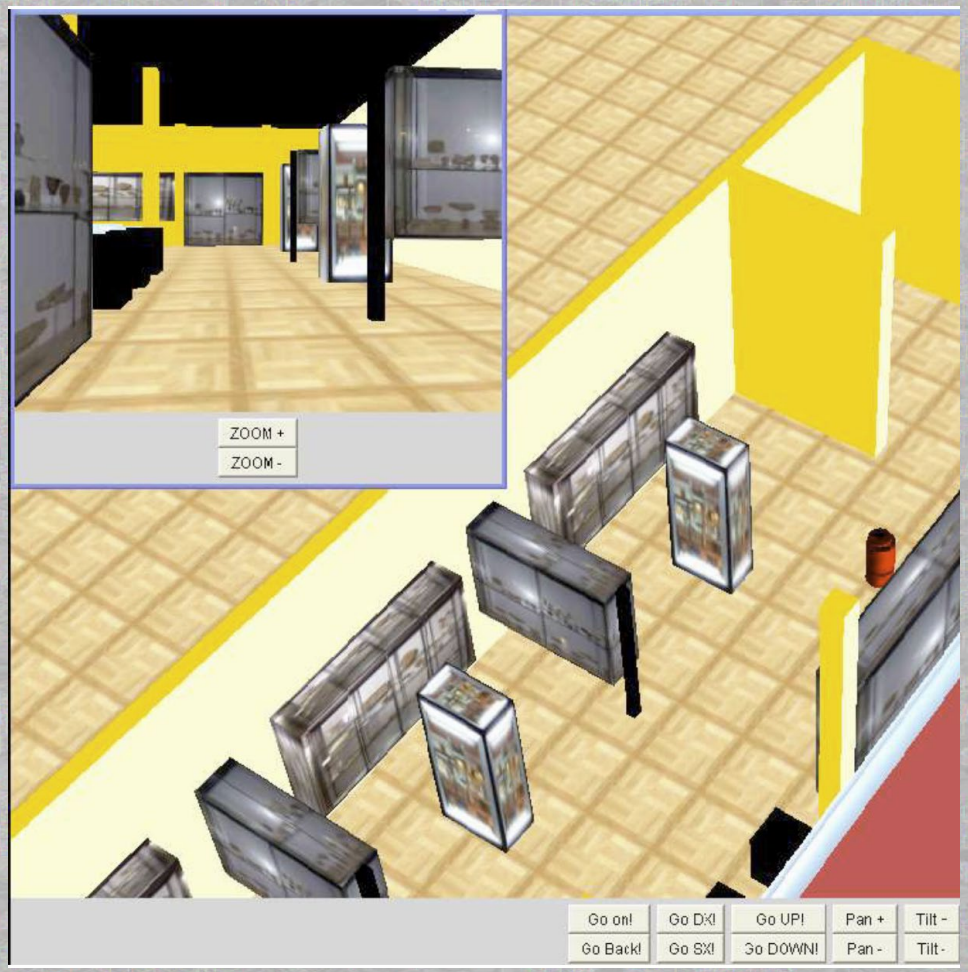
MC3 Machine Consciousness

- Long tradition of describing the structure of consciousness from a first-person perspective.
- For example, Husserl and Merleau-Ponty.
- Can create computer models of conscious experiences in a machine.
- This is MC3 machine consciousness.

Imagination with CRONOS and SIMONS



Cicerobot


ZOOM +
ZOOM -
Go on!
Go Back!
Go DX!
Go SX!
Go UP!
Go DOWN!
Pan +
Pan -
Tilt -
Tilt -

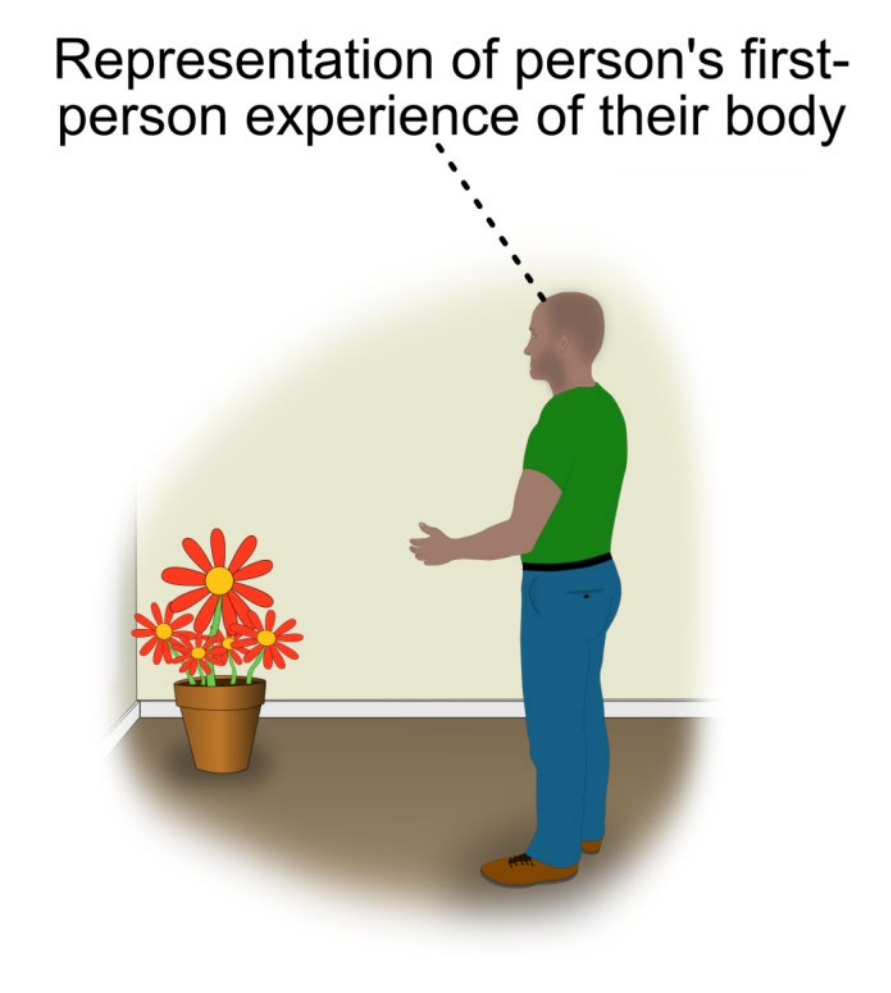
Types of Machine Consciousness

- **MC1**. Machines with the same external behaviour as conscious humans.
- **MC2**. Computer models of the correlates of consciousness.
- **MC3**. Computer models of consciousness.
- **MC4**. Machines that really have conscious experiences.

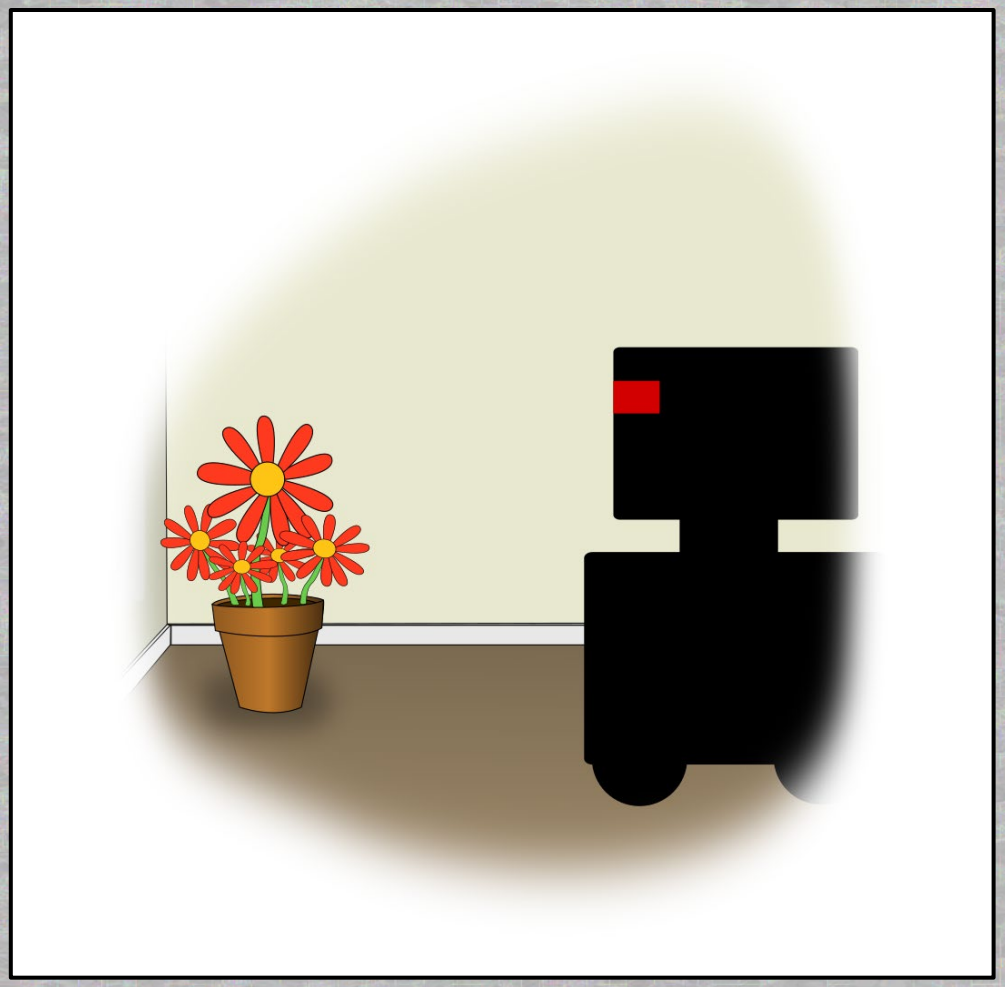
MC4 Machine Consciousness

- A physical robot is MC4 conscious if it is associated with a bubble of experience.
- Its bubble of experience will contain something analogous to our colours, smells etc.

Human Consciousness



Conscious Machine (MC4)

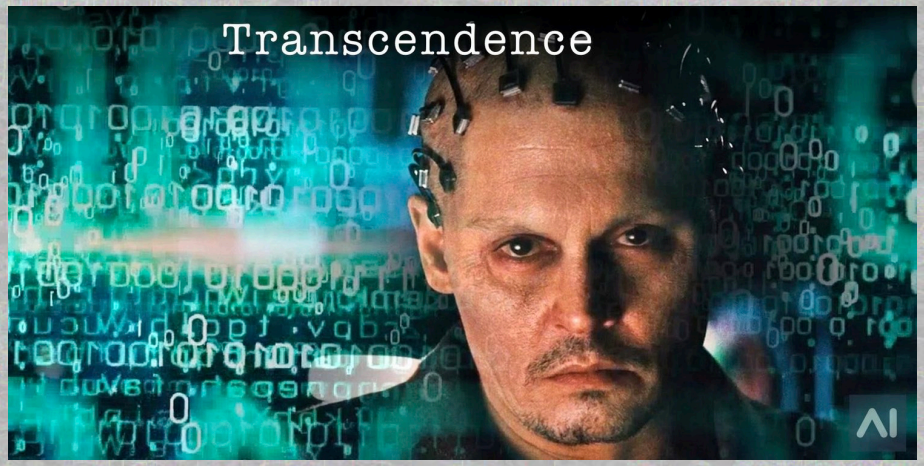


Significance of Research on MC4 Machine Consciousness

- Ethical issues.
- Curiosity.
- We want to achieve immortality.
- Medical applications.
- Helps us to develop general scientific theories of human consciousness.

MC4 Conscious Machines

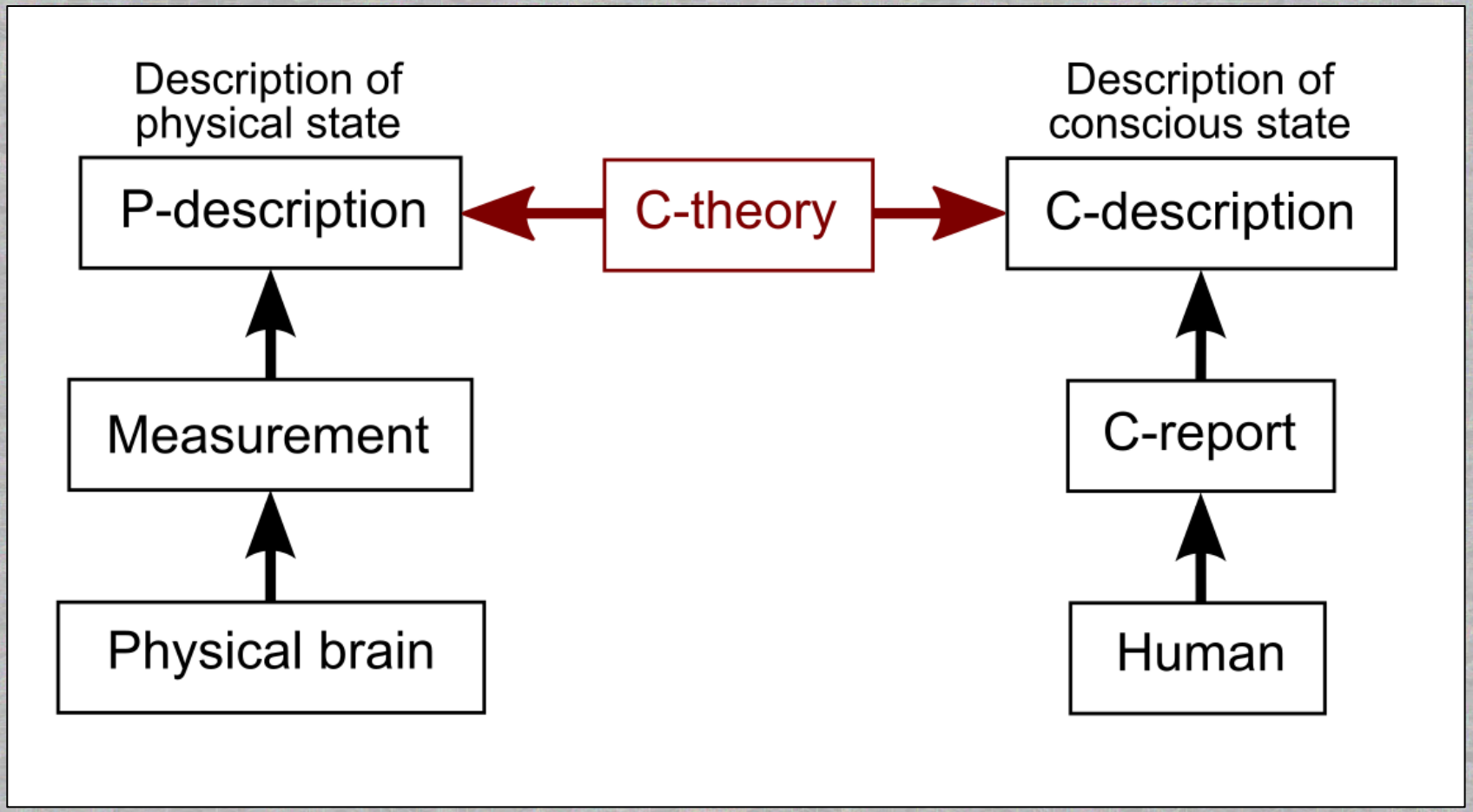
MC4 Consciousness Transfer / Uploading



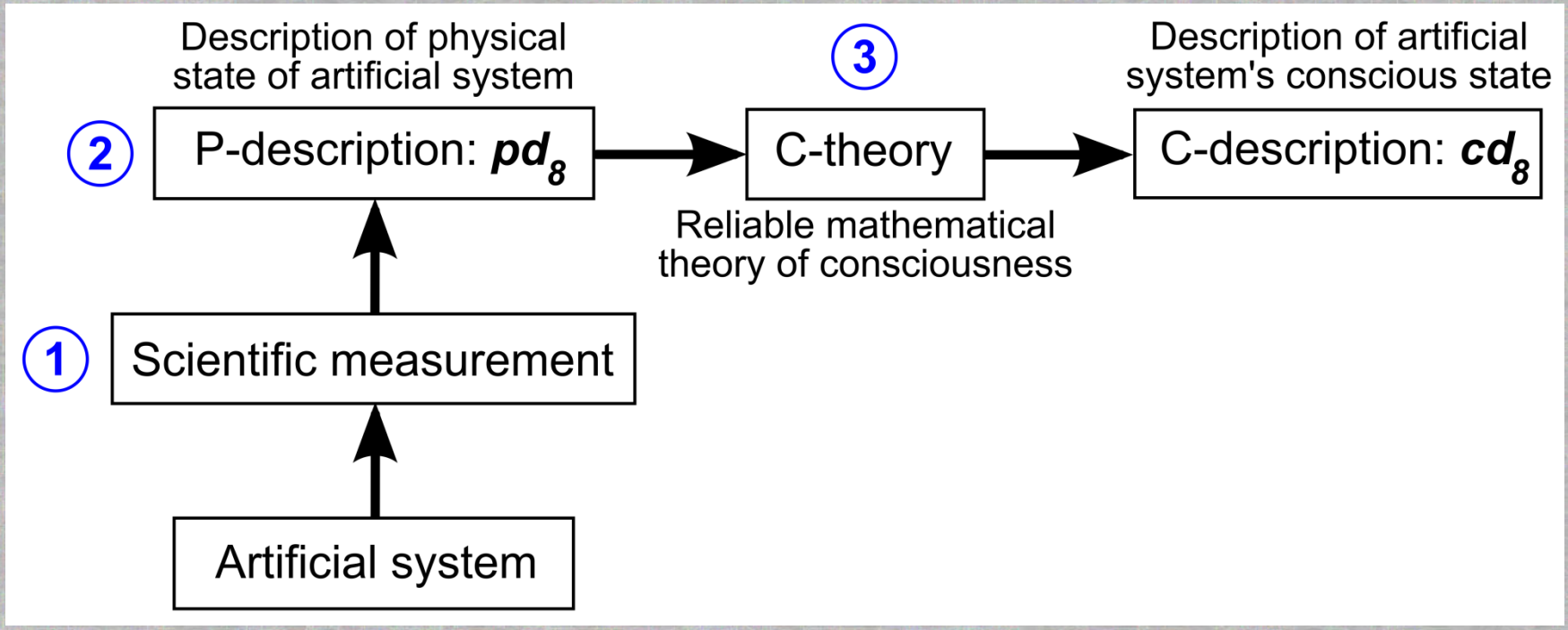
Solving MC4 Machine Consciousness

- Mathematical theories of consciousness can solve problem of MC4 machine consciousness.
- Use c-theory to:
  - Make plausible deductions about the consciousness of a machine.
  - Build machines that are associated with specific conscious states.

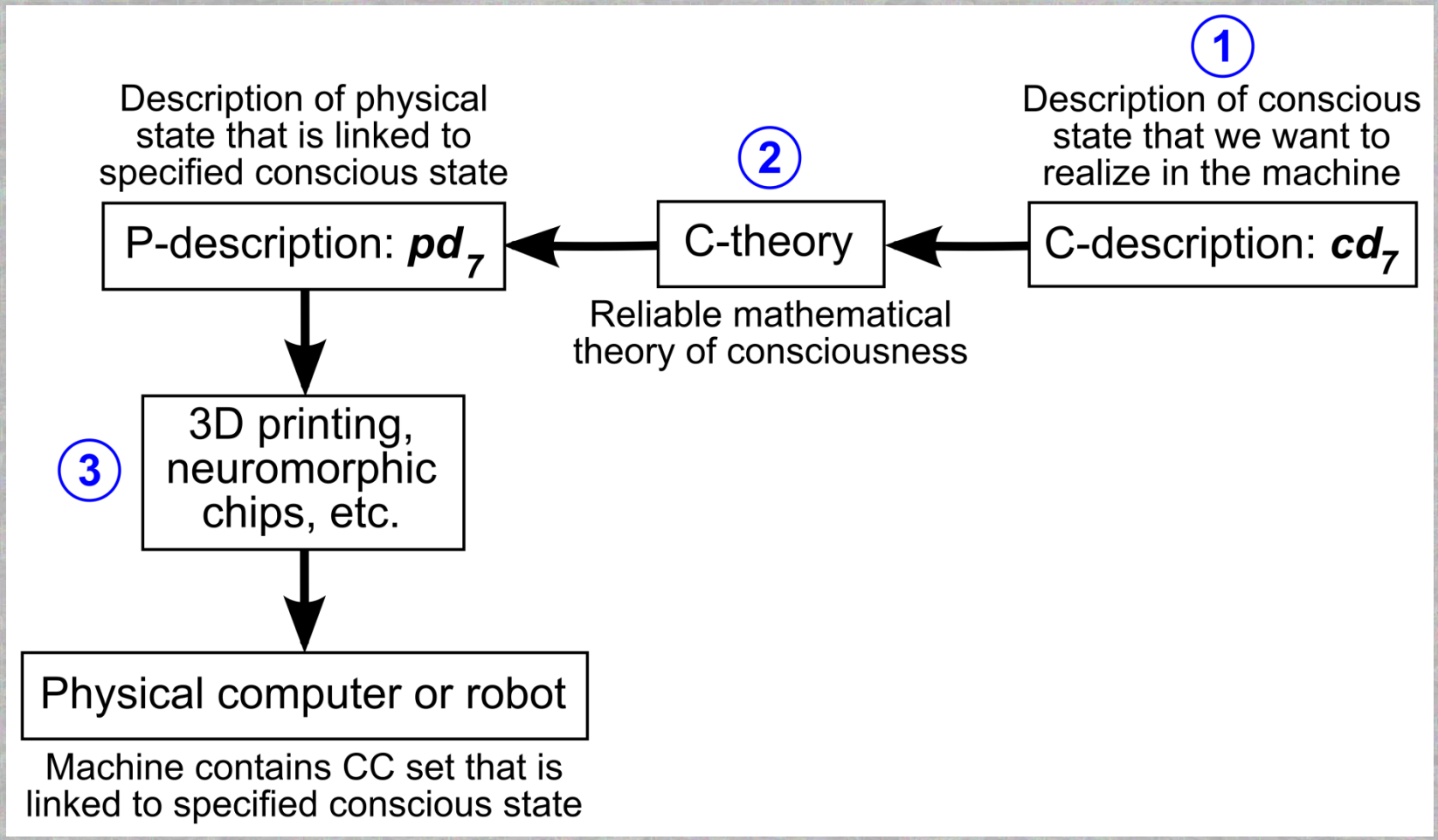
Mathematical Theory of Consciousness



Deducing the MC4 Consciousness of a Machine



Building a MC4 Conscious Machine



Summary

- Models of the neural correlates of consciousness are developed by neuroscientists to understand the relationship between consciousness and the brain.
- Models of the correlates of consciousness and models of consciousness are built by AI researchers to produce more intelligent machines and help us to understand how we can analyze machines for consciousness.

CONCLUSION

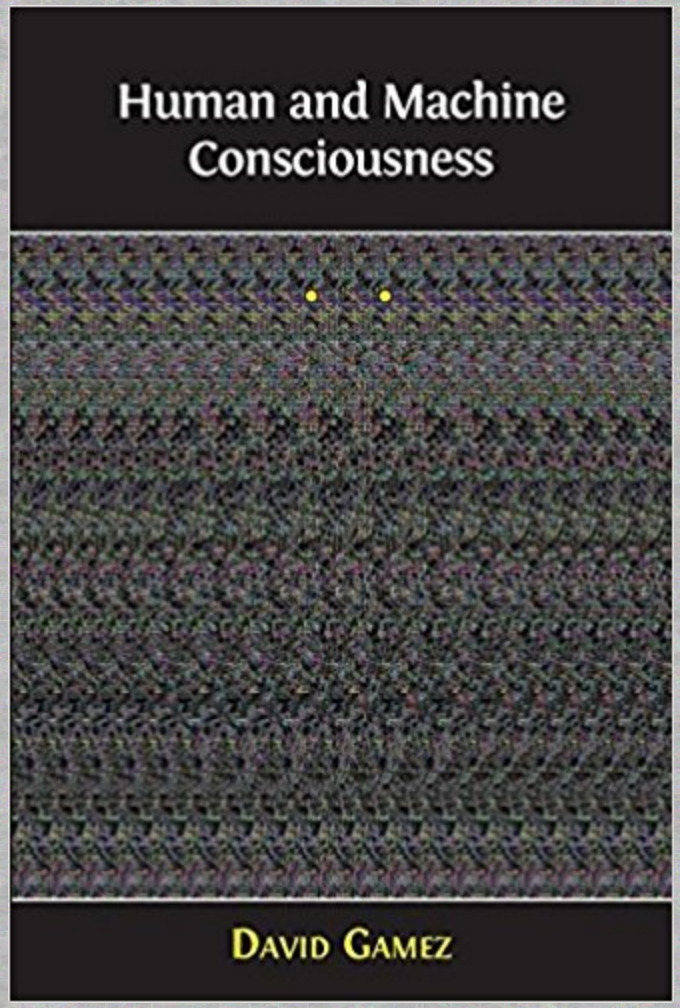
Conclusion

- Modern concepts of consciousness and the physical world *co-evolved* – we cannot understand one without the other.
- Imagination and thought experiments are of limited use for studying the relationship between consciousness and the physical world.
- We should use scientific methods to discover mathematical relationships between measurements of consciousness and measurements of the physical world.

Conclusion

- Models of the neural correlates of consciousness can help us to understand how potential neural correlates of consciousness are implemented in the brain.
- Can use models of the correlates of consciousness and models of consciousness to build more intelligent machines.
- Mathematical theories of consciousness might eventually be able to determine whether artificial systems are MC4 conscious.

More Information



- Read for free, download and/or purchase at: https://www.openbookpublishers.com/product/545.
- Website with papers: www.davidgamez.eu.

Questions?

- Website: www.davidgamez.eu.
- Book: www.openbookpublishers.com/product/545
- Contact: david@davidgamez.eu.