The Hard Problem of Consciousness:
300 Years On

David Chalmers

Explaining Consciousness

- How can we explain consciousness?

- Can consciousness be explained in physical terms?

- Can there be a science of consciousness?

The Easy Problems of Consciousness

- The easy problems: explain the objective functions associated with consciousness

  - perceptual discrimination

  - integration of information

  - control of behavior

  - verbal report

- One can explain these in physical terms by specifying a mechanism that performs the function

The Hard Problem

- Explain why and how physical processes are associated with subjective experience?

    - Why is there something it is like to be me?

    - Why is it like this?

- This is not a question about objective functions. It's a further question.

Outline

1. History of the Hard Problem

2. The Key Argument

3. Materialist Responses

4. Nonreductive Theories

5. Machine Consciousness

6. The Science of Consciousness

History of the Hard Problem

Brihaspati (600BC)

Brihaspati (600BC)

"Earth, fire, air, and water, are the ultimate existents. Their combination is called the body, senses, and objects. Consciousness arises out of these ultimate existents, as the power to intoxicate arises out of fermenting ingredients."
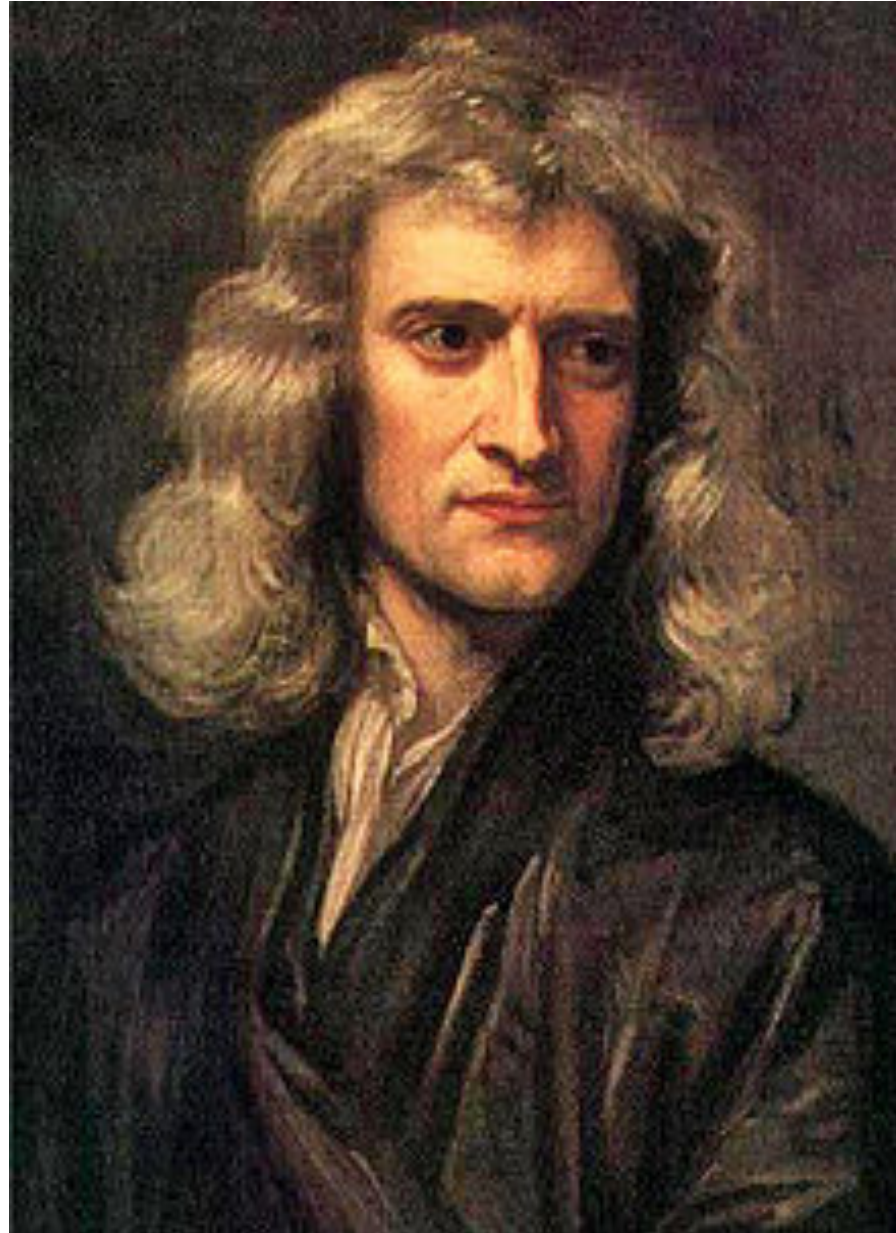
Galen (150AD)

Galen (150AD)

"A single body capable of sensation cannot be produced from many which are incapable of sensation. Sensation certainly is of a different genus than shape, weight, or hardness, which belong to the atoms, or than the others that belong to fire, air, earth, and water. Consequently, the body that is capable of sensation cannot be constituted either from atoms or from fire, air, earth, and water."

René Descartes?

Isaac Newton

Newton (1672)

- "to determine by what modes or actions light produceth in our minds the phantasm of colour is not so easie." (letter to Henry Oldenburg)

Newton (1672)

- "But, to determine more absolutely, what light is, after what manner refracted, and by what modes or actions it produceth in our minds the phantasms of colours, is not so easie." (letter to Henry Oldenburg)
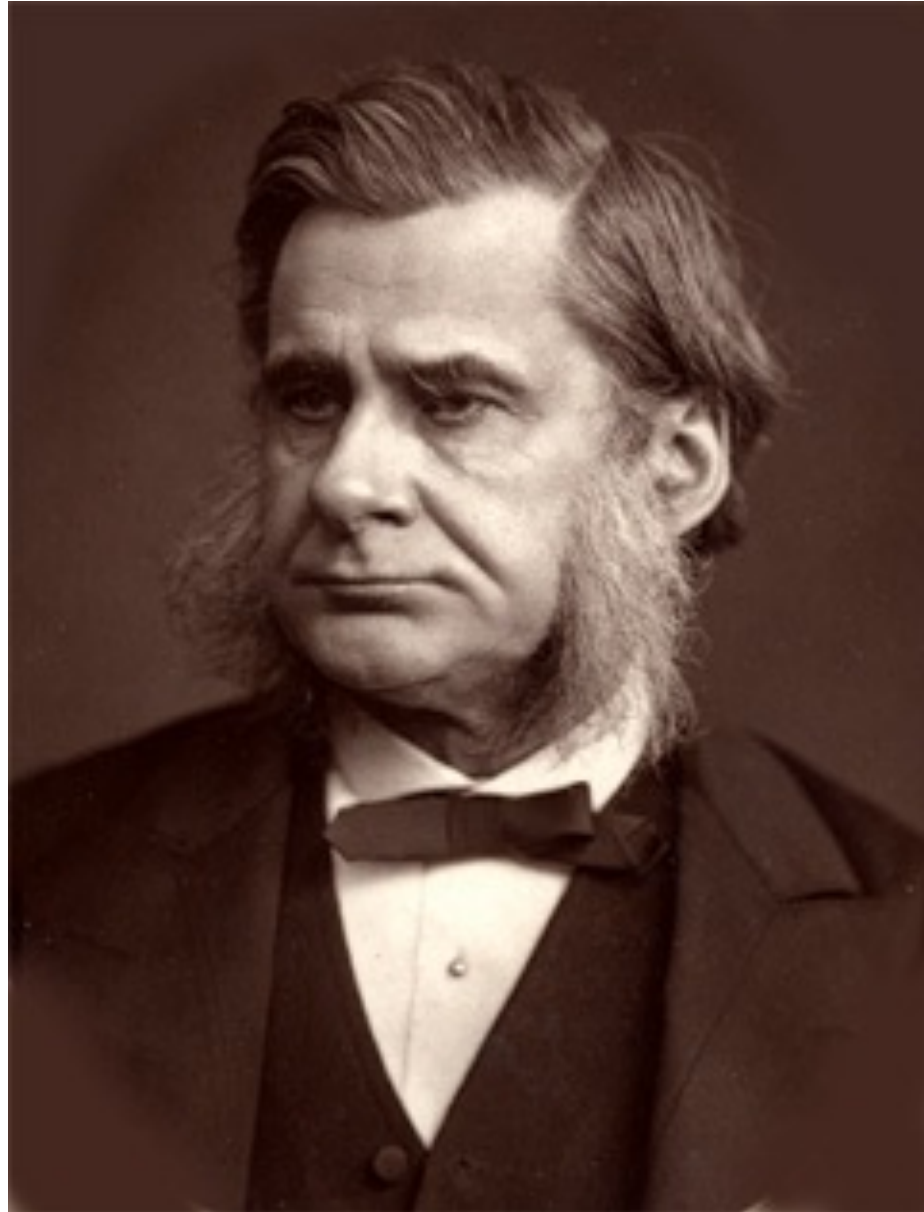
Gottfriend Wilhelm Leibniz

Leibniz (1714)

- "Moreover, it must be confessed that perception and that which depends upon it are inexplicable on mechanical grounds, that is to say, by means of figures and motions. And supposing there were a machine, so constructed as to think, feel, and have perception, it might be conceived as increased in size, while keeping the same proportions, so that one might go into it as into a mill. That being so, we should, on examining its interior, find only parts which work one upon another, and never anything by which to explain a perception. (*Monadology*)
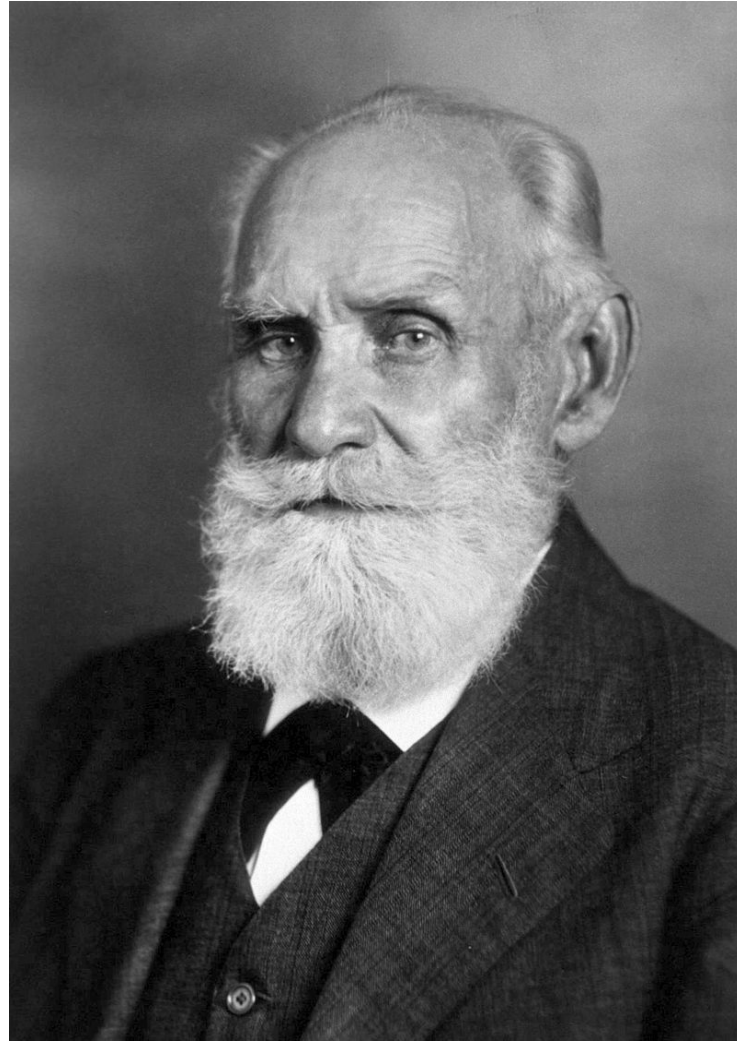
Thomas Huxley

Huxley (1866)

- "How it is that anything so remarkable as a state of consciousness comes about as a result of irritating nerve tissue, is just as unaccountable as the appearance of the Djin when Aladdin rubbed his lamp." (*The Elements of Physiology and Hygiene)*

Ivan Pavlov

Pavlov (1923)

"Allow me to take this opportunity to express in a few words how we represent physiologically what we call "consciousness" and "conscious." Certainly I will not discuss this question from the philosophical point of view, i.e., I shall not touch on the problem of how the brain substance creates subjective phenomena, etc. I shall only endeavour to answer provisionally what kind of physiological phenomena, what sort of nervous processes, proceed in the hemispheres of the brain when we say we are "conscious" and speak of our "conscious" activity."   ["Twenty Years Experience of Objective Studies of Animal Higher Neural Activity"]

Pavlov (1923)

- "Philosophical" question "How does a matter of brain produce subjective phenomenon?"

- "What physiological phenomena, what neural processes do exist in large hemispheres, when we say we are conscious of ourselves, when our conscious activity takes place?" "Twenty Years Experience of Objective Studies of Animal Higher Neural Activity" [Pavlov I. Complete Works. 2nd ed. V. 3(1). Moscow: AS USSR, 1951. P. 247.]

Thomas Nagel

Nagel (1974)

- "Consciousness is what makes the mind-body problem really intractable. … Without consciousness the mind-body problem would be much less interesting. With consciousness it seems hopeless." ("What is it like to be a bat?")

David Chalmers

Chalmers (1994)

- Hard problem vs easy problems of consciousness

  - (1) Catchy name

  - (2) Distinction between problems *of consciousness*.

  - (3) The labels encapsulate an argument.

Argument

- (1) Purely physical explanations explain only the easy problems (objective functions)

- (2) Explaining consciousness requires more than explaining the easy problems

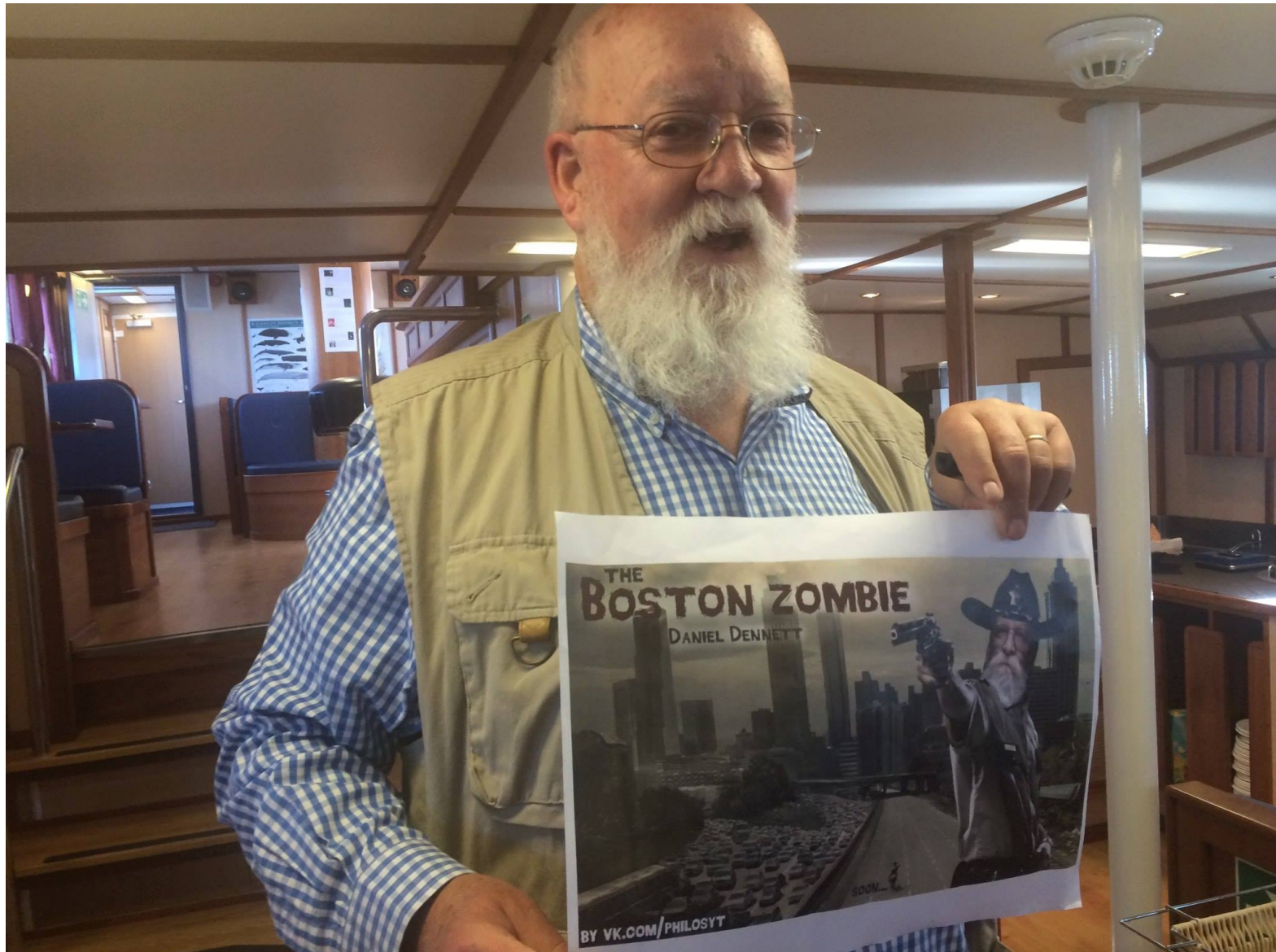- So (3) No purely physical explanation can explain consciousness.

What's Happened Since?

- 1. Materialist responses

- 2. Nonreductive theories

- 3. Science of consciousness

Materialist Responses

- 1. No hard problem: Explaining the objective functions explains everything. [Type-A materialism]

- 2. The hard problem involves an epistemic gap, not an ontological gap. [Type-B materialism]

- 3. Enrich the microphysical to incorporate (proto)consciousness [panpsychism, neutral monism]

Type-A materialism

Type-A materialism

- No hard problem: Explaining the objective functions explains everything that needs to be explained. [Dennett, …]

- Either (i) consciousness doesn't exist, or (ii) explaining the functions explains consciousness.

Type-A materialism

- Type-A materialism is an important view, but it has been surprisingly unpopular and under-developed over the last 20 years.

  - Even archreductionists tend to acknowledge the hard problem: Crick, Koch, Kurzweil, Pinker, …

- I think the type-A view deserves developing

Illusionism about Consciousness

Illusionism about Consciousness

- Consciousness is an illusion (Dan Dennett, Nicholas Humphrey, Keith Frankish, Derk Pereboom).

- We can functionally explain the things we say about consciousness. Once we have done this, we have explained the illusion of consciousness.

- Challenge: give a good functional explanation, and show that this is all that needs explaining.

Type-B Materialism

Type-B Materialism

- The hard problem involves an epistemic gap, not an ontological gap. [Balog, Block, Carruthers, Hill, Papineau, Tye, …]

- There's a gap between our concepts of the physical and our concepts of consciousness, but consciousness itself is physical all the same.

- Problem: This view seems to require that our concepts of consciousness are themselves physically inexplicable.

Nonreductive Theories

- Consciousness is a fundamental property, not reducible to physical properties but connected to them by fundamental laws.

  - Dualism: epiphenomenalism or interactionism

  - Pan(proto)psychism: panpsychism or panprotopsychism (neutral monism)

  - Idealism

Dualism

Dualism

- Dualist theories face the problem of interaction: either no causal role for consciousness, or finding a role within physics.

- Leading approach: a role for consciousness in collapsing quantum wave functions? (Stapp, Hodgson, Chalmers/McQueen).

Dualism

- Dualist theories face the problem of interaction:

  - epiphenomenalism: no causal role for consciousness (counterintuitive?).

  - interactionism: consciousness affects physics (unscientific?)

Dualism and Quantum Mechanics

- Leading interactionist approach: a role for consciousness in collapsing quantum wave functions? (Stapp, Hodgson, Chalmers/McQueen).

Panpsychism

Pan(proto)psychism

- Consciousness or protoconsciousness is present at the microphysical level (Strawson, Rosenberg, Seager, Goff, Coleman, Tononi, Koch, Hameroff/Penrose)

- Russellian panpsychism: Consciousness serves as the intrinsic nature underlying physical structure, and is the causal basis for microphysical action.

- Constitutive panpsychism: Microphysical consciousness adds up to our macroconsciousness.

- Together: yields a causal role for consciousness consistent with physics and integrated with it.

The Combination Problem

- The combination problem for pan(proto)psychism: how do microexperiences add up to macroexperience?

  - subject combination problem, quality combination problem, structure combination problem

  - no new fundamental laws of combination!

- No-one has a good solution to this problem yet.

Idealism

Idealism

- The physical world exists only in the minds of observers (Berkeley, Hoffman)

- Problem: We need something outside our experience to explain the regularities in our experience.

- Leads back to either panpsychism or dualism.

Science of Consciousness

- How does all this connect to the science of consciousness?

- How can a scientist contribute to the hard problem of consciousness?

Recent History

- In the last twenty years, numerous scientific theories of consciousness have been put forward

- Some relatively reductionist:

  - e.g. neuronal global workspace theory

- Some relatively nonreductionist

  - e.g. information integration theory

Fundamental Theories

- A number of researchers have developed quasi-empirical theories of consciousness take consciousness to be fundamental and postulate fundamental laws
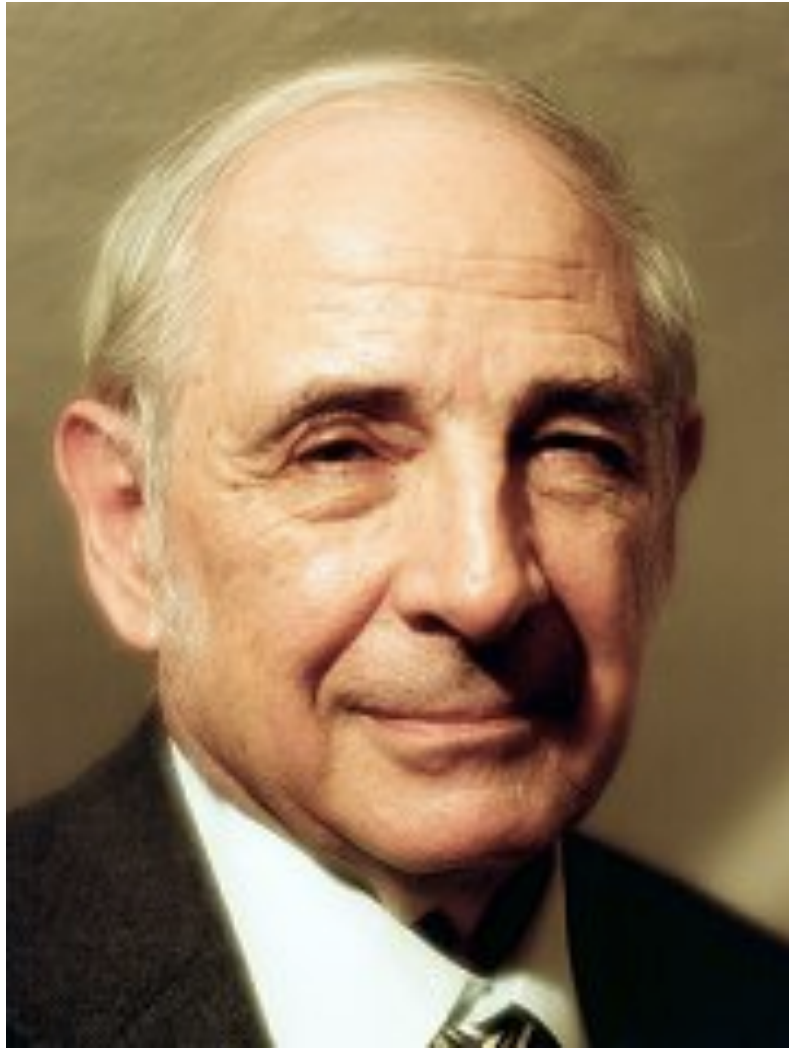
What Fundamental Theory?

- What should be the key notion in a fundamental theory of consciousness?

- One speculation: information!

Machine Consciousness

- Can a machine be conscious?

- We don't know how.

- But we don't know how brains can be conscious either!

- Are computers worse off than brains?

Thought Experiments

- Thought-experiments on machine consciousness

  - John Searle, "Minds, brains, and programs" (the Chinese room), 1981

  - Ned Block, "Troubles with functionalism" (the Chinese national), 1978

  - Anatoly Dneprov, "The game" (the Portuguese stadium), 1961

Anatoly Dneprov

И Г Р А

Анатолий ДНЕПРОВ   РАССКАЗ   Рисунки И. УШАКОВА

Это была, как сказал профессор Зарубин, «математическая игра настоящей воды».

Участвовать в ней предложили нескольким делегатам Всесоюзного съезда молодых математиков, и, к всеобщему удивлению, молодыми оказались все тысяча четыреста человек. Игра происходила на большой арене стадиона имени Ленина.

— Учтите, игра будет продолжаться часа три-четыре. Так что наберитесь терпения. Если кто не выдержит — все пропало! — предупредил Иван Клочко, молодой украинский логист. Ему Зарубин поручил вести всю организационную работу, которая выглядела очень странно.

— Запомните номер вашей команды. Вам присваивается номер 10. Каждого участника вы сами занумеруйте порядковыми числами в двоичной системе, 1-й, 10-й, 11-й и т. д. — говорил Иван главе представителей от Российской Федерации.

Так он подходил ко всем делегациям, сообщал им условный индекс и разъяснял порядок нумерации участников.

На организованную игру ушла вся суббота, и сбор был назначен на 9 утра в воскресенье. Мне казалось, что ни один человек, который принял участие в этой организационной неразберихе, не забудет ее до конца жизни.

Ровно в 9 утра все мы собрались на стадионе. Там уже прохаживался профессор Зарубин, его ассистент Семен Данилович Рябов и Иван Клочко.

Зеленое поле стадиона было расчерчено оранжевыми лентами на квадраты и прямоугольники. В каждой фигуре стояли небольшие деревянные тумбочки, на голубой поверхности которых мелом был написан номер. Все мы решили, на прав, ожидая что-то, что будет дальше.

Профессор Зарубин куда-то исчез, и вскоре мы услышали его голос, раздававшийся по радио по всему стадиону.

— Группа участников с индексом 1011, займите прямоугольную площадку в восточном крыле стадиона. Располагайтесь шеренгами и в каждой друг за другом по порядку выдвинутой руки, в порядке порядкового номера. Сами человека в шеренге, глубина строя — шесть человек.

— Группа с индексом 101, займите поле у южных трибун. Также располагайтесь шеренгами друг за другом, на расстоянии вытянутой руки, в порядке возрастания номеров. Группа с индексом...

В течение пятнадцати минут Зарубин подробно инструктировал все группы участников, кому, где и как расположиться. Как только профессор называл индекс группы, молодежь вскакивала и стайкой бежала на указанный участок стадиона.

— А сидеть можно!? — крикнул кто-то.

Через несколько секунд веселый голос Зарубина сообщил:

— Можно! Главное, строго соблюдайте тот порядок, который я вам указал.

Я принадлежу к так называемой специальной команде. Мне и моим товарищам предстояло расположиться между отдельными полями и, как объяснил Клочко, «быть связными между командами».

Когда построение было закончено и стадион принял вид, как будто полтора тысячи юношей и девушек собрались для выполнения коллективных гимнастических упражнений, снова послышался голос профессора Зарубина.

— Теперь слушайте правила игры.

— Начиная с северной трибуны, вверно с товарища Сагирова, будут передаваться числа в двоичной системе исчисления. Например, индекс один-ноль-ноль-один. Товарищ Сагиров сообщит эту цифру соседу справа, если она приходится с цифрой один, и соседу слева, если она приходится с цифрой ноль.

Если в числе будут последовательно два единицы или два нуля, то он должен сообщить это число соседу, следующему за его стоящей следующей шеренге. Каждый, получив от своего соседа числовое сообщение, должен прибавить к нему свой порядковый номер и, в зависимости от результата сообщить его соседу. Кроме того, если группа имеет индекс...

И так далее.

Правила игры были повторены три раза, и когда на вопрос: «Понятно!?» весь стадион хором ответил: «Понятно!», Зарубин сказал:

— Тогда начали.

Стол между группами «110» и «1001», в виде его ассистент Зарубина, Семен Данилович — кто-то товарищ делегации Грузинской ССР, Нажмиров, для них была необходима еще важно-то особые инструкции.

Игра началась ровно в десять утра.

В конце, как началась с северной трибуны, говоря участников начали поворачиваться то вправо, то влево. По полям и эти движения распространялись...

39

Dneprov, "The Game"

"When the layout was complete the stadium looked like a large gym with fourteen hundred of young people inside going to do exercise. Then again came the Professor's voice: "Here are the rules. Binary numbers will be given to comrade Sagirov from the northern stand. For instance, "one-zero-zero-one". If the first digit is "one", comrade Sagirov is to pass the number to the person on his right, whereas all numbers starting with "zero" shall go to the person on his left.""

Dneprov, "The Game"

"This is a sentence in Portuguese. I don't think you can guess what it means. However, it was you who yesterday made a perfect Russian translation. To save you the trouble of guessing, I want to explain what the game actually was. In short, we can call it a Computing Machine game. Each one of you was either a memory cell, a total mechanism, a time-delay line or a simple switch."

Dneprov, "The Game"

"Remember that part of Turing's article where he said that to find out whether machines are able to think, you have to become a machine. Experts in cybernetics believe that the only way to prove that machines can think is to turn yourself into a machine and examine your thinking process. Hence, yesterday we spent four hours operating like a machine."

Dneprov, "The Game"

"If you, being structural elements of some logical pattern, had no idea of what you were doing, then can we really argue about any thoughts of electronic devices made of different parts which are deemed incapable of any thinking even by the most fervent followers of the electronic brain concept? ... I think our game gave us the right answer to the question "Can machines think?" We've proven that even the most perfect simulation of machine thinking is not the thinking process itself which is the higher form of motion of the living matter."
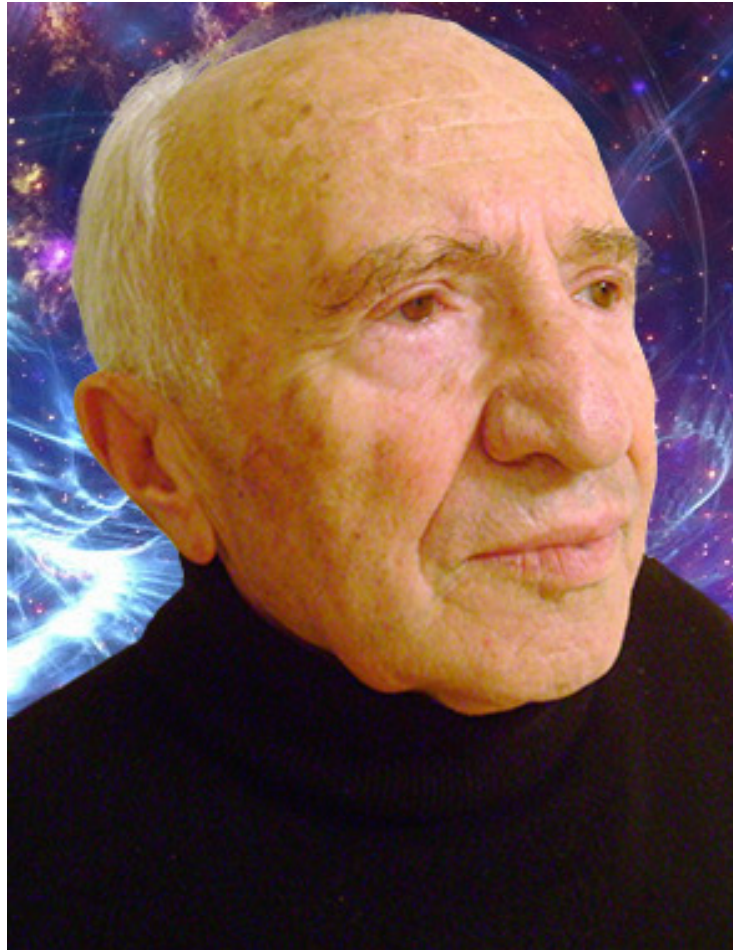
Systems Reply

- Systems Reply: The consciousness of the stadium system is not identical to the consciousness of any of the people.

- If you gradually replace my neurons by tiny people, I'll still be conscious of Portuguese, but the people won't.

Moral

- Moral: We must distinguish the consciousness of a machine from the consciousness of any components.

- What matters is the information processed by the system as a whole.
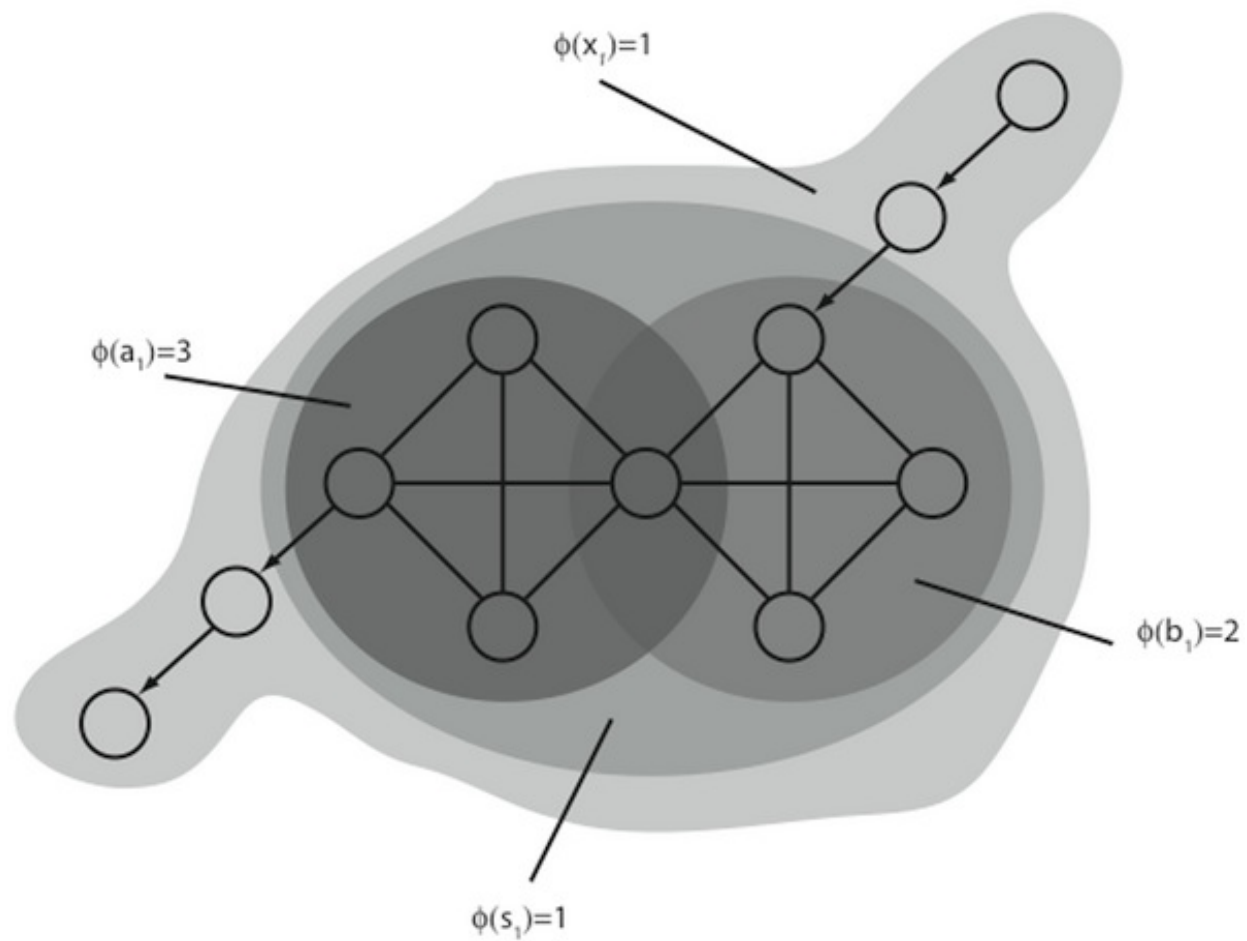
Informational Approaches

David Dubrovsky

- "Every phenomenon of consciousness is a piece of information, since it is intentional and represents something. Since any information is necessarily embodied in its material bearer, in the given case the bearer is a particular neurological process. This, in principle, provides an answer to the question of a necessary connection between the 'mental' and the 'physical'."

Double-Aspect Theory of Information (Chalmers)

- Information has two aspects: a physical aspect and a phenomenal aspect.

- The fundamental psychophysical laws should be formulated in terms of information.

Integrated Information Theory (Giulio Tononi)

$$ei(x_1; P) = -\sum_{i=1}^{k}\sum_{\mu_0^{(i)}} p(\mu_0^{(i)}|x_1) \log p(\mu_0^{(i)}|\mu_1^{(i)}) - H(X_0|x_1)$$

$$\Phi(x_1) = \min_{P} \frac{ei(x_1; P)}{v_P}$$

Integrated Information Theory

- consciousness <-> integrated information

- phi: a measure of information integration

  - high phi <-> high consciousness

  - low phi <-> low consciousness

Information and The Metaphysics of Consciousness

- The informational approach can be combined with various different metaphysics of consciousness.

Type-A Materialist IIT

- Consciousness is wholly explainable in terms of the dynamics of information

  - Explain integration dynamically, nothing else needs explaining

- Dubrovsky: there is no explanatory gap.

Type-B Materialist Version

- Consciousness is identical to and reducible to integrated information

  - a primitive theoretical identity, as with classic mind-brain identity theory?

Epiphenomenalist Version

- Integrated information causes consciousness. There's a closed dynamics of information and a psychophysical laws linking that dynamics to consciousness.

- Consciousness doesn't play a causal role in the dynamics, so it is epiphenomenal.

- So consciousness is epiphenomenal?

Interactionist Version (Chalmers and McQueen)

- Bidirectional psychophysical laws:

  - High phi causes consciousness

  - Consciousness collapses the quantum wave function.

- IIIT: Integrated information interactionist theory!

- Testable in principle.

Panpsychist Version

- The physical world is a world of information, with consciousness as its intrinsic nature.

- Information is everywhere, so consciousness is everywhere.

- To solve the combination problem: we need to understand the principles of composition for information.

My view

My view

- I divide my credence about 50-50 between pan(proto)psychism and property dualism.

- If we can solve the combination problem, then pan(proto)psychism.

- If one can rigorously make sense of a causal role for nonphysical consciousness in quantum mechanics, then property dualism.

Two Paths

- Currently I'm actively pursuing both paths.

- Today: the combination problem seems so strong that I tentatively favor interactionism, perhaps via quantum interactionist IIT.

- Tomorrow: who knows?

Conclusion

- A theory of consciousness is an empirical project: find a fundamental theory that best fits the scientific data.

- But it's also a project heavily constrained by philosophical reasoning.

- A project for scientists and philosophers working together for the next 300 years.