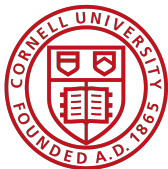


# Welcome to Cartesian Theater

## An Introduction to Metaphysics of Consciousness

Minakami Yuki

Cornell University, Connective Media



Based on Jaegwon Kim and Daniel Dennett's paper.

# Table of Contents

1. Background
2. Physicalism & Dualism
3. Arguments against Physicalism
4. Arguments for Physicalism
5. Russellian Monism
6. Conclusion

# Background

## Why Cognitive Science Is Not Enough

All sorts of mental phenomena have yielded to scientific investigation in recent years, but consciousness has stubbornly resisted.

- **World Knot:** Why does a state of consciousness come about as a result of initiating nerve tissue?
- **Access Consciousness:** A mental state is access-conscious if its content is processed in a way that it can be used to control reasoning and behavior.

# The Hard Problems

The really hard problem of consciousness is the problem of **experience**.

- When we think and perceive, there is a whirl of information processing, but there is also a **subjective** aspect.
- As Nagel has put it, there is **something it is like** to be a conscious organism.

Looking at a red rose, smelling it, feeling the thorns on its stem: these experiences all feel a certain way as well. We will call these feels '**phenomenal feels**'.

Philosophers have presented a range of arguments to show that phenomenal feels cannot be reduced to anything purely physical.

- Knowledge Argument
- Philosophy Zombies
- What is it like to be a bat?

# Explanatory Gap

It seems that no matter how much information we are given about conscious experience in physical, functional or intentional terms, that information won't really explain why our experiences should feel to us **the way that they do**.

Philosophers in the last decade or so have fleshed out a powerful essence to these arguments by invoking **qualia**.

**Qualia:** We use the term 'qualia' to refer to the introspectively accessible, phenomenal aspects of our mental lives.

- In this broad sense of the term, it is difficult to deny that there are qualia.
- Qualia are at the very heart of the mind-body problem.

Why and how such-and-such objective physical changes generate so-and-so subjective feeling or any subjective feeling at all?

# Why Qualia Matters?

The existence of qualia may seem to some to mark the threshold for states or creatures that are really conscious.

- If an organism senses and responds in apt ways to its world but lacks such qualia, then it might count as conscious at best in a loose and less than literal sense.
- Or so at least it would seem to those who take qualitative consciousness in the “what it is like” sense to be significant.



# Possible Worlds

Most of us believe that things, as a whole, needn't have been just as they are. Rather, things might have been different in countless ways, both trivial and profound.

- In any case, no matter how things had gone they would still have been part of a single, maximally inclusive, world.
- Intuitively, then, the actual world of which our immediate situation is a part is only one among many possible worlds.

# Metaphysical Level

Modal metaphysics concerns the metaphysical underpinning of our modal statements. These are statements about what is possible or what is necessarily so.

- Cognitive science can't deal with the consciousness problems in a set of possible worlds.
- The essence of consciousness is usually intended as a metaphysical thesis.

# Physicalism & Dualism

## Physicalism

**Physicalism:** Everything is fundamentally physical.

- Physicalists don't deny that the world might contain many items that at first glance don't seem physical.
- But they insist nevertheless that at the end of the day such items are physical, or at least bear an important relation to the physical.

# A More Explicit Statement

## Supervenience

The idea of supervenience has its origins in meta-ethics but was imported into philosophy of mind mainly by Davidson.

**Supervenience:** Every domain of entities supervenes with metaphysical necessity on the physical domain.

- Necessarily, there isn't any mental difference without a physical difference.

This is the **standard formulation** of Supervenience.

# Necessity Physicalism

$S_1$  : Physicalism is true at a possible world  $w$  iff every property instantiated at  $w$  is necessitated by a physical property.

- Physicalism concerns some possible worlds and not others, offered by the necessity to an arbitrary possible world.
- Suppose we focus on an arbitrary possible world, say the actual world.

# Supervenience Physicalism

$S_2$  : Physicalism is true at a possible world  $w$  iff any world which is a physical duplicate of  $w$  is a duplicate of  $w$  simpliciter.

- It will be useful to have a name for physicalism so defined, so let us call it supervenience physicalism.
- A duplicate simpliciter: a world that is exactly like ours in all respects.

**Dualism:** The mental and the physical are, in some sense, radically different kinds of things.

- Common sense tells us that there are physical bodies, one could say that materialist monism is the 'default option'.
- Dualism, therefore, tends to start from the assumption of the reality of the physical world.

**Cartesian Dualism:** A normal human being involves two substances, one a body and the other a person.

- The latter is a purely mental substance that can be defined in terms of consciousness alone, as Descartes claimed.
- The mind is not just a collection of thoughts, but is *that which* thinks, an immaterial substance over and above its immaterial states.



**Property Dualism:** There are two essentially different kinds of property out in the world.

In the case of mind, property dualism is defended by those who argue that the qualitative nature of consciousness is not merely another way of categorizing states of the brain or of behaviour, but a genuinely emergent phenomenon.

# Arguments against Physicalism

Now turn to the truth question:

Is physicalism (as we have interpreted it so far) true?

Before providing an answer to this question, I will discuss the arguments for supposing that physicalism is not true.

Then I will consider the arguments for physicalism.

# Knowledge Argument (KA)

Mary is a brilliant scientist who is, for whatever reason, forced to investigate the world from a black and white room via a black and white television monitor.

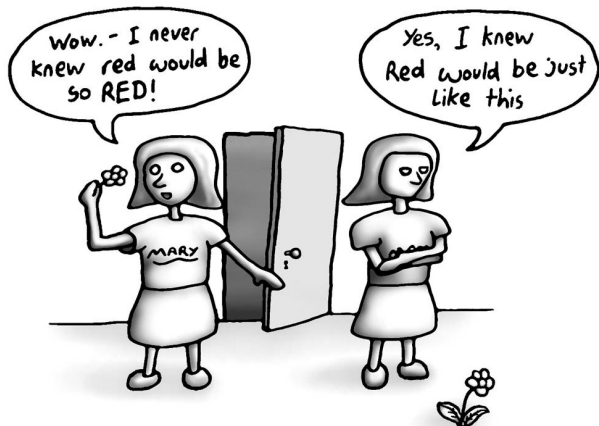
- She specializes in the neurophysiology of vision and acquires, let us suppose, all the physical information there is to obtain about what goes on when we see ripe tomatoes, or the sky, and use terms like 'red', 'blue', and so on.

# Mary the Scientist

She discovers, for example, just which wavelength combinations from the sky stimulate the retina, and exactly how this produces via the central nervous system the contraction of the vocal chords and expulsion of air from the lungs that results in the uttering of the sentence 'The sky is blue'....

- When Mary is finally released from the black and white room, she sees the sky and shouts: "AHA, *that* is blue."
- **The Key Question:** At that moment, does Mary acquire any new knowledge?

# Mary the Scientist



# Key Question

It seems just obvious that she will learn something about the world and our visual experience of it.

- Then is it inescapable that her previous knowledge was incomplete?
- But she had all the physical information.
- *Ergo* there is more to have than that, and physicalism is false.

# Knowledge Argument (KA)

(P<sub>1</sub>) Mary has complete *physical knowledge* concerning facts about human color vision before her release.

(P<sub>2</sub>) But there is *some kind of knowledge* concerning facts about human color vision that she does not have before her release.

(C) There is *some kind of knowledge* concerning facts about human color vision that is non-physical knowledge.

## Explicit formulation of KA

(P<sub>1</sub>) Mary has complete *physical knowledge* concerning facts about human color vision before her release.

(C<sub>1</sub>) Mary knows all the physical facts about human color vision before her release.

(P<sub>2</sub>) There is some (kind of) knowledge concerning facts about human color vision that Mary does not have before her release.

(C<sub>2</sub>) There are some facts about human color vision that Mary does not know before her release.



# Explicit formulation of KA

Therefore, from (C<sub>1</sub>) and (C<sub>2</sub>):

(C<sub>3</sub>) There are non-physical facts about human color vision.

(P<sub>3</sub>) If physicalism is true, then physical information captures everything.

(C<sub>4</sub>) Therefore, physicalism is false.

# Conceivability Argument (CA)

The argument that I have just given is a version of what is usually called a conceivability or modal argument against physicalism. It entails a metaphysical premise:

(PP) A situation is possible in our cognitive imagination iff it is possible metaphysically (in the objective world independent of our cognition).

In a nutshell, whatever is conceivable is possible.

**Phenomenal Zombie:** An organism exactly like a human being but which completely lacks phenomenal consciousness.

- From a phenomenal zombie's perspective, there is nothing it is like to be in their mental states.
- A phenomenal zombie may still have mental states in the functionalist sense.

# Explicit formulation of CA

(P<sub>1</sub>) It is conceivable that my zombie twin exists.

With (PP):

(P<sub>2</sub>) It is possible that my zombie twin exists.

(P<sub>3</sub>) I have phenomenal consciousness but my zombie twin doesn't.

(P<sub>4</sub>) The zombie twin and I are identical in every physical respect.

(C) Therefore, there is something that is not physical; physicalism is false.

As usually presented, KA says that it is possible to know all the physical facts without knowing the psychological facts, and that in consequence physicalism is false.

- KA can be thought of as a conceivability argument or modal argument too.
- It is simply that here we are interested in whether it is possible for a person to know all the physical facts without knowing all the facts.

## Rebuttal to (PP)

Even if a zombie world is *conceivable*, that does not establish that it is *possible* in the way that matters.

- Conceivability is an epistemic notion, while possibility is a metaphysical one.
- Given psychophysical identities, it is an 'a posteriori' fact that any physical duplicate of our world is exactly like ours in respect of positive facts about sensory states.

## A Posteriori Physicalism

If physicalism is true, then there is a conditional  $C^*$ , which is necessarily true. According to a posteriori physicalism, this conditional is necessary and a posteriori.

( $C^*$ ) It is not a priori that there is someone identical to me in respect of all physical truths but different with respect to some psychological truth.

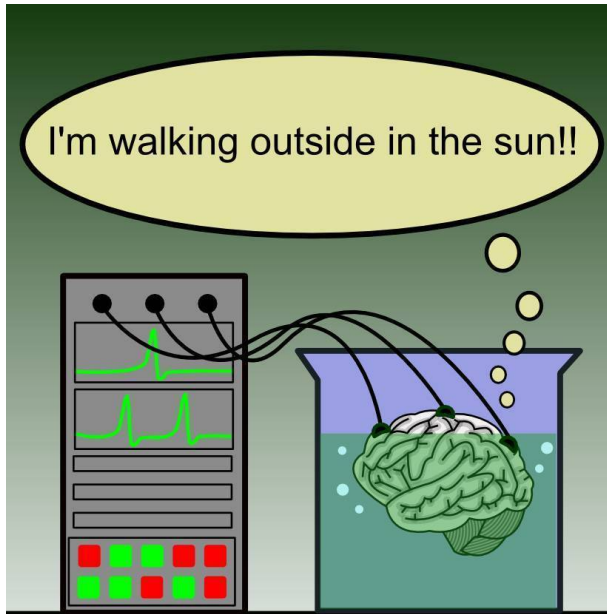
On this view, what 'it is conceivable that  $p$ ' means is 'it is not a priori that not  $p$ .'

So in particular, what 'it is conceivable that water is not  $H_2O$ ' means is that 'it is not a priori that water is not  $H_2O$ .'

Now in the argument we are considering we are in effect already talking about the negation of the proposition that physicalists are attached to.



## A Popular Question (Off Topic)



# Brain in a Vat(BIV)

**Brain-in-a-Vat:** A given person is a disembodied brain living in a vat of nutrients. The nerve endings of the brain are connected to a supercomputer.

(P<sub>1</sub>) If an agent cannot distinguish between two situations, then that agent cannot know which of those two situations obtains.

(P<sub>2</sub>) We cannot distinguish between reality and BIV.

---

(C<sub>1</sub>) Therefore, we cannot know whether reality or BIV obtains.

(C<sub>2</sub>) Since there are external objects in reality but not in BIV, we cannot know whether there are any external objects either.

# Moore's proof of an External World

(P<sub>1</sub>) Here (holding up one's left hand) is one hand.

(P<sub>2</sub>) Here (holding up one's right hand) is another.

---

(C<sub>1</sub>) Therefore, there are at least two hands.

(C<sub>2</sub>) Therefore, there are at least two things to be met with in space.

(C) There is an external world.

**Common Sense Realism:** Our ordinary common-sense view of the world is largely correct.

Moore then goes to argue that this demonstration of his hands was a 'perfectly rigorous' proof of the existence of external objects.

- For its premises certainly entail its conclusion and they are things which he then knew to be true.
- Moore's avowed aim is to prove his knowledge of the existence of an external world.

# Moore's Strategy

It turns out that his strategy here is more subtle; he wants to argue that we get our understanding of knowledge primarily through straightforward cases of this kind, and thus that skeptical arguments are self-undermining.

- On the one hand, they rely on general principles about the limits of knowledge and thus assume some understanding of knowledge.
- On the other hand, they undermine this understanding by implying that there are no such straightforward cases of it.

# Arguments for Physicalism

## The Impressionistic Argument

**The Epistemological Thesis:** We should believe the world-view of science whatever that view happens to be.

Physicalism is the thesis that everything is necessitated by physical facts, where this means something like 'the facts described by contemporary or near contemporary physics.'

**The Interpretative Thesis:** Physicalism is somehow implicit in the world-view of modern science.

A more powerful argument uses principles about causal reasoning.

# Causal Argument

**Causal Closure:** The system of physical causes and effects in the world is causally closed in the sense that every effect in that system has a cause from within that system.

- (1) For all physical events  $e$ , if there is an event  $e^*$  such that  $e^*$  causes  $e$ , then  $e^*$  is a physical event.

**Causal Efficacy:** Mental events cause physical effects.

- (2) For any mental event  $e$ , there is a physical event  $e^*$  such that  $e$  causes  $e^*$ .

# Exclusion Principle (EP)

**EP:** (3) For all events  $e$  and  $e^*$ , if  $e$  causes  $e^*$ , then there is no event  $e^{**}$  such that

- (a)  $e^{**}$  is distinct from  $e$ .
- (b)  $e^{**}$  causes  $e^*$ .

The suggestion that (1)–(3) entail that every mental event is a physical event seems reasonable on the surface.



# The Property Version

(1\*) For all physical events  $e$ , if there is an event  $e^*$  such that  $e^*$  causes  $e$ , then

- (a)  $e^*$  is a physical event.
- (b) There is a physical property  $F$  such that  $F$  is causally efficacious in  $e^*$ 's causing  $e$ .

(2\*) For all mental events  $e$ , there is a physical event  $e^*$  such that  $e$  causes  $e^*$  and there is a mental property  $F$  such that  $F$  is causally efficacious in  $e$ 's causing  $e^*$ .

# The Property Version

(3\*) For all events  $e$  and  $e^*$ , if  $e$  causes  $e^*$ , and if there is a property  $F$  which is causally efficacious in  $e$ 's causing  $e^*$ , then there is no property  $G$  such that

- (a)  $G$  is distinct from  $F$ .
- (b)  $G$  is causally efficacious in  $e$ 's causing  $e^*$ .

Not just for the conclusion that every mental event is a physical property but that every mental property is a physical property.

Therefore, physicalism is true.

# Explanatory Gap

*How* can consciousness exist?

One may concern whether the phenomenal and qualitative properties of a conscious creature's mind can be a *priori* deduced from a description of the neural properties of its brain processes.

Our present inability to see any way of closing the gap may exert some pull on our intuitions, but it may simply reflect the limits of our current theorizing rather than an unbridgeable in principle barrier.

# Explanatory Gap

If one's premises concern physical or neural facts, then one will need some bridge principles or links that connect such facts with facts about consciousness.

They would probably not allow us to see how or why those connections hold, and thus they would fall short of fully explaining how consciousness exists.

**Russellian Monism:** A single set of properties underlies both consciousness and the most basic entities posited by physics.

- The theory is named for Bertrand Russell, whose views about consciousness and its place in nature were informed by a structuralist conception of theoretical physics.

# Structuralism in Physics

**Structuralism:** Physical theory only characterizes its basic entities relationally, in terms of their causal and other relations to other entities.

- Basic particles, for instance, are largely characterized in terms of their propensity to interact with other particles.
- Their mass and charge is specified, to be sure, but all that a specification of mass ultimately comes to is a propensity to be accelerated by certain forces, and so on.

The picture of the physical world that this yields is that of a giant causal flux, but the picture tells us nothing about what all this causation relates.

- Suppose structuralism about physics is explicated such that structural properties are construed as extrinsic properties or relatively intrinsic properties.
- In that case, the thesis is consistent with the claim that physical theory characterizes its basic entities partly in terms of its intrinsic properties.

# Wave-Particle Duality

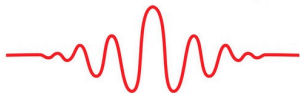
**Wave-Particle Duality:** Every particle or quantum entity may be described as either a particle or a wave.

- It expresses the inability of the classical concepts "particle" or "wave" to fully describe the behaviour of quantum-scale objects.
- Quantum field theory generates an entirely new concept of fields that exist in ordinary space-time, causality still being visualizable.



# Wave-Particle Duality

## Wave-Particle Duality



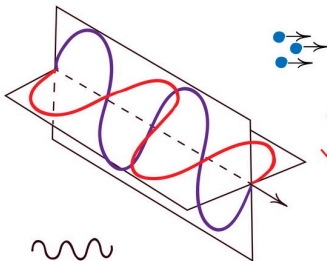
wave



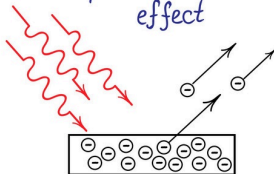
$$c = 299\,792\,458 \text{ m/s}$$



particle



photoelectric effect



# Something Similar

Traditional physicalism has the virtue of allowing consciousness to be integrated into physical causation.

- But this comes at the cost of disregarding or distorting the distinctive features of consciousness.

Traditional dualism avoids that cost.

- But by separating consciousness from the physical world, traditional dualism prevents consciousness from being integrated into physical causation.

Russellian monism combines theoretical virtues of traditional dualism with those of traditional physicalism, while avoiding their main drawbacks.

- Russellian monism provides a plausible response to anti-materialist arguments such as the conceivability argument and the knowledge argument.
- Those arguments might show that structure alone does not fully determine (metaphysically necessitate) the nature or even the existence of consciousness.

# Conclusion

Reviewing the previous content, we learned

- the basic concepts of physicalism
- anti-physicalism arguments
- causal argument
- the flaws of physicalism and the explanatory gap
- the basic concepts of Russellian Monism

I am very glad to be able to taste the wonder of philosophy of mind with all of you!

# References



## Facing up to the problem of consciousness.

Chalmers, David.

*Journal of Consciousness Studies*. Vol 2, No. 3, 1995, pp. 200-19.



## Physicalism, or Something near Enough.

Kim, Jaegwon.

*Philosophical Quarterly*. Vol 56, No. 223, 2006, pp. 306-10.



## What is Russellian Monism?

Alter, Torin & Nagasawa, Yujin.

*Journal of Consciousness Studies*. Vol 19, No. 9, 2012, pp. 67-95.

# Thanks and Questions

Thank you for listening!

Welcome to ask any questions!

Minakami Yuki

[minnakamiyuki@gmail.com](mailto:minnakamiyuki@gmail.com)

## Selected Questions

In order to avoid silence, I give some questions for discussion at the end.

1. Does physicalism deny free will?
2. Do you think it's possible for AI to be conscious?
3. Which metaphysics of consciousness do you support and why?
4. Supervenience comes from metaethics, so can you give an example of moral supervenience?
5. We usually think that there is a big difference in methods of thinking between men and women (in a traditional sense). How to explain this with philosophy of mind?