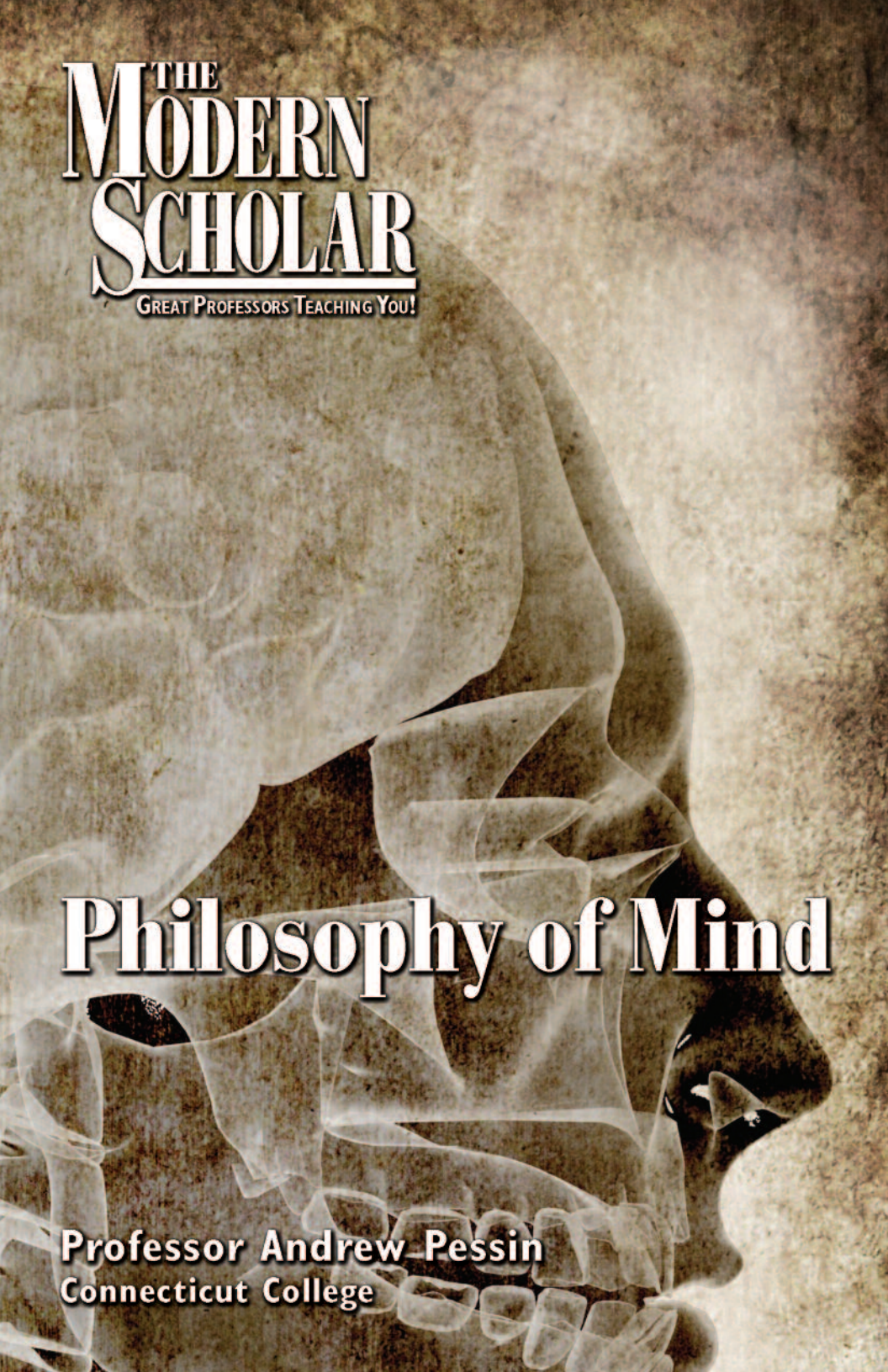

THE MODERN SCHOLAR
GREAT PROFESSORS TEACHING YOU!
Philosophy of Mind
Professor Andrew Pessin
Connecticut College

Philosophy of Mind

Andrew Pessin
Connecticut College

Philosophy of Mind

Professor Andrew Pessin

Executive Editor

Donna F. Carnahan

RECORDING

Producer - David Markowitz

Director - Ian McCulloch

Podcast Host - Gretta Cohn

COURSE GUIDE

Editor - James Gallagher

Design - Edward White

Cover image: © Shutterstock.com

#UT160 ISBN: 978-1-4407-7251-1

Course Syllabus

Philosophy of Mind

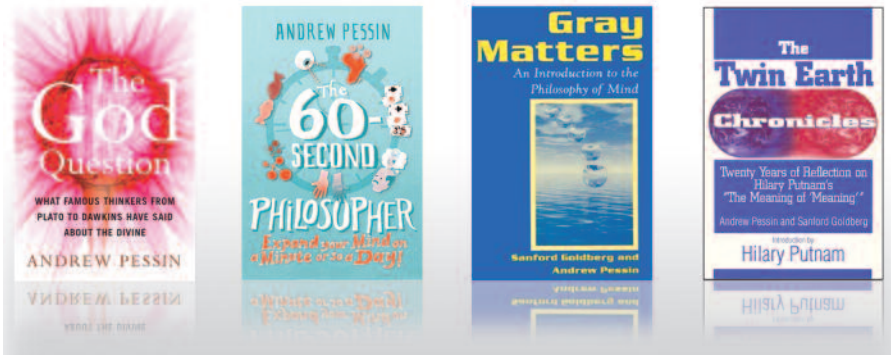
Photo courtesy of Andrew Pessin

About Your Professor

Andrew Pessin

Andrew Pessin is a professor of philosophy and chairman of the Philosophy Department at Connecticut College in New London. He studied at Yale and Columbia, has spent two decades teaching liberal arts to undergraduates at Columbia University, the College of William and Mary, Kenyon College, Wesleyan University, and Connecticut College, and has given talks about philosophy to many non-philosophical audiences. Professor Pessin's teaching and lecturing has had the goal of making philosophy entertaining, accessible, and, most of all, fun. He has appeared several times on the *Late Show with David Letterman* as "The Genius."

Professor Pessin is the author of two best-selling philosophy books for general audiences: *The God Question: What Famous Thinkers from Plato to Dawkins Have Said About the Divine* (Oneworld Publications, 2009) and *The 60-Second Philosopher: Expand Your Mind on a Minute or So a Day!* (Oneworld Publications, 2009). He is the coauthor (with Sanford Goldberg) of two other books: *Gray Matters: An Introduction to the Philosophy of Mind* (M.E. Sharpe, 1997) and *The Twin Earth Chronicles: Twenty Years of Reflection on Hilary Putnam's "the Meaning of 'Meaning'"* (M.E. Sharpe, 1996). Professor Pessin has also published many academic articles and in his spare time composes and performs songs about philosophy. For more information, visit his website at **www.andrewpessin.com**.



Cover images: © Oneworld Publications and M.E. Sharp

Introduction

Look briefly at a bright light, then close your eyes and focus on what you see. You should detect an "afterimage," which starts out almost as bright as the original light and then begins to fade. But now consider the question, "Where exactly is this (say) greenish image you see?" It's not outside your head, since your eyes are closed. Nor is it inside your head: there's nothing green on the inside of your eyelid, after all, nor on your retina, nor is there anything green inside your brain, which is just a lot of wet, mushy gray-and-white stuff. So if the greenish image is neither outside nor inside your head, where is it?

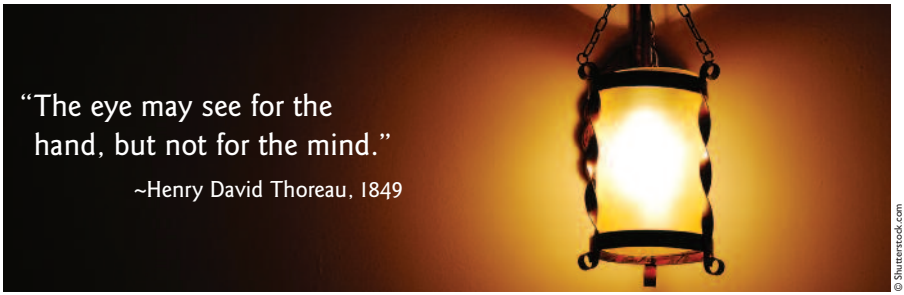
Welcome to the Philosophy of Mind.

What is the nature of the mind, and how does it relate to the body? Could consciousness and thought and emotions just be physical states of a brain? What does it mean, exactly, to have a "thought," and what precisely is an emotion? Can computers ever truly think, or can they at best merely simulate thinking? This course examines various philosophical attempts to make sense of the mind, and of matter, and of the relationship between them. Along the way we will explore some of the most famous thought experiments in contemporary philosophy and meet such unusual creatures as thinking computers, people with inverted color vision, residents of a place called "Twin Earth," and zombies.

Lecture 1

Mind and Body

The Suggested Reading for this lecture is Kirk Ludwig's "The Mind-Body Problem: An Overview" in *The Blackwell Guide to Philosophy of Mind,* edited by Stephen P. Stich and Ted A. Warfield, chapter 1, pp. 1–46.



Introducing the Mind

It was suggested in the introduction that "afterimages" create a mystery. Where are they? Where did they go? The most natural answer to these questions, of course, is this: afterimages are "in the mind." But this is no ultimate answer at all, because it promptly generates the next round of questions. If afterimages are not located in the brain, but they *are* "in minds," then minds themselves must not be located in the brain. But then where, exactly, are they? Further, if minds are not located anywhere in space, then they have no spatial character or dimensions: no length, breadth, depth—and lacking these, they would lack volume. But if they lack volume, then nothing could literally be "inside" them. And most importantly, if minds are not themselves spatial in character, then how could they have any relationship, such as a causal relationship, to the brains that clearly have a lot to do with causing our behaviors—and which *are* located in physical space?

These are some of the key questions we'll be exploring in this course.

Our first task is to get a little clearer on just what we mean when we talk about "minds."

Examples of the "Mental"

First of all, we might give some paradigm examples of what the mind is *not*, namely, examples of purely physical or material things: rocks, trees, water, stars, atoms, and molecules, for instance.

In contrast, paradigm examples of mental things, that is, things pertaining to the mind, might include sensations, thoughts, beliefs, desires, emotions or feelings, intelligence, character, and personality.

Sensations

These come in various forms: pains, itches, tickles, afterimages, seeing a round red patch (when looking at an apple), hearing screeching car tires, or feeling nausea. What sensations all share is that they have a certain "qualitative nature": there is a certain "way they feel" or "there is something that it is like to have them." This qualitative nature of our experiences, which philosophers refer to as "qualia," is rather hard or impossible to describe in words.

Thoughts, Beliefs, Desires

These are grouped together as examples of what philosophers often call "propositional attitudes." In these cases the subject is said to have an "attitude" toward a proposition, where the proposition specifies the "content" of the state. We might observe (for example) that Fred is thinking that it's time to buy a new car; the subject is "Fred," and he has the attitude of "thinking" toward the proposition "It's time to buy a new car." But there are many varieties of propositional attitudes: in addition to thinking, believing, and desiring, one may "hope that" something, "fear that" something, "be outraged that" something, or "desire that" something.

Unlike sensations with their hard-to-put-into-words qualia, propositional attitudes seem to have an intimate relationship to language. It's often suggested that propositional attitudes are also different from sensations in simply lacking qualia altogether.

Feelings/Emotions

Some examples of feelings and emotions include anger, joy, sadness, depression, and remorse. As with propositional attitudes, these are often attributed using propositions: we say someone is angry that something is the case, or that he feels depressed that something is the case. Yet unlike propositional attitudes they *do* seem to have a qualia associated with them: there is something it feels like to feel angry, or sad, or depressed, and so on. And while their qualia make them similar to sensations, there is also a difference: sensations seem rather directly to be caused by the things that cause them in a way quite different from how emotions are caused.

Intelligence, Character, Personality

These are often seen as "derivative" traits, being ultimately related to more basic mental states. To be "intelligent," for example, is to be able to quickly form the right beliefs; to be morally good is to have the tendency to form the right sorts of desires, and so on.

Our primary focus will be on what seem to be the most basic mental phenomena, namely, sensations and propositional attitudes.

The Mark of the Mental

If all of the above are examples of the "mental," can we specify just what it is about them that makes them count as "mental"? Are there distinguishing "marks of the mental" that set mental things apart from purely physical things?

Philosophers suggest various criteria for the mental, each covering various cases, but perhaps no single one captures everything. In looking at these we'll focus on a "qualia" example—the mental pain caused by, say, some physical nerve damage in your tooth—but most of what follows applies to propositional attitudes as well.

Four Interrelated "Knowledge Criteria" for Distinguishing the Mental from the Physical

Direct (or Immediate) Knowledge

You know you have the pain "directly or immediately," that is, not based on any evidence or reasoning or inference. You just simply *feel* it. Contrast this with your knowledge of the physical nerve damage itself. How do you know there is some physical problem with your physical tooth? Well, the painful feeling itself suggests to you that something is wrong with your tooth—it is your evidence that something is wrong with your tooth. But then you don't know about the physical damage "directly" or "immediately": you only know about it by feeling your pain, and then doing some reasoning or inference about the cause of the pain.

Privacy, or First-Person Privilege

Mental states seem "private" to the person having them, meaning, at least, that that person has a special kind of access to his own mental states that no other person can have to those same states.

So, developing the first criterion: I have direct access to my mental states, that is, I can know them directly and immediately, but everyone else only has indirect access to my mental states. I know my toothache pain by feeling it, but the dentist knows that I am feeling toothache pain only indirectly, on the basis of evidence. Physical things, to the contrary, do not support this kind of privacy or privileged access.

Infallibility

One's knowledge of one's own current mental states, it is suggested, is *infallible*: you cannot be mistaken about whether you're experiencing them. In contrast, any belief about a purely physical occurrence can be in error.

For a dramatic example, consider the "phantom limb syndrome." Some people, after suffering the amputation of a limb, report sensations coming as if from the limb. So imagine someone suffering from this feels an itch in his missing leg. Clearly he cannot be wrong that he is feeling an itch, but he may be wrong in believing the physical cause of the itch is "in his leg."

Self-Intimacy

Mental states are what philosophers call "self-intimating," namely, it's impossible to have or be in a mental state without knowing that you are. So, for example, you can't have a pain you're unaware of. In contrast, there can be all sorts of purely physical things going on in us of which we are unaware, even the kind of things that normally produce our pains.

Are the "Knowledge Criteria" for Distinguishing the Mental from the Physical Successful?

These "knowledge criteria" do, no doubt, suggest very important properties traditionally ascribed to the mind and to things mental, and it's clear that many mental states do display them much of the time. But right now, at least, we're seeking criteria such that *all* mental phenomena fulfill them, and *no* physical phenomena do, without exception. And the "knowledge criteria" do not succeed in providing that. For it seems that perhaps not all mental phenomena display the properties we've been discussing, and perhaps not all the time.

So, for example, we often speak of "unconscious" beliefs, desires, emotions—allegedly mental or psychological states we are in even without being aware of them. We recognize that we can often be wrong in various ways about our mental states. And it's often quite unclear whether we have a privileged access to many of our propositional attitudes, and even to our sensations.

What we might conclude, then, is that the knowledge criteria provide a useful "first blush" way of distinguishing what's "mental" from what's "physical"—but that they simply are not perfectly exhaustive.

Two Metaphysical Criteria for Distinguishing the Mental from the Physical

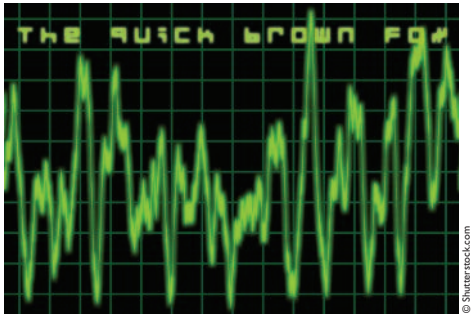
Spatiality

This is the property of "being spatial": having spatial dimensions and being located in space. According to René Descartes, everything physical displays or possesses spatiality while nothing mental does. In recent decades, there has arisen much to call into question just how complete this criterion truly is. Contemporary physics, for example, may not quite endorse the idea that all physical things are truly spatial in nature. In the reverse direction, there are things that lack spatial character and yet are not "mental" in nature. For example, philosophers speak of "abstract objects" such as numbers.

Intentionality

"Intentionality" is the property of "aboutness": something has this property insofar as it is capable of having an object, or "being about," something. A physical piece of wood isn't "about" anything at all, but the mental thought Fred has about buying a new car *does* have an object, *is* about something, namely, that car. Because propositional attitudes have such objects, they are often called "intentional states" as well.

But do all mental states in fact display intentionality? That is, do they have an object they are about? Many think that sensations don't seem to have a content in this way. Worse, there seem to be genuinely physical things that *do* display intentionality. Linguistic objects such as words and sentences have it, for example: the sentences you are hearing are purely physical objects, they are sounds, or vibrating air molecules, and yet insofar as they are meaningful there is something they are about.



Intentionality is clearly an important property, but more has to be said—and will be, in later lectures—before it will serve as a clear way of demarcating the physical from the mental.

Conclusion

We began by exploring various paradigm examples of physical and mental things. We then asked just what it is, exactly, that distinguishes the mental from the physical. We examined a variety of candidate criteria. Although all are of some use, none of them are criteria such that all mental things have them, and no physical things, or vice versa. What remains unclear, in other words, is whether we can obtain a single, unifying concept of the mental—a single thing that all mental states share, and all physical things lack.

Nevertheless, what emerges is that what we called "qualia" and "intentionality" are extremely important phenomena that will provide much of the subject matter of our course.

FOR GREATER UNDERSTANDING

Questions

1. Are there important examples of "mental" entities or phenomena beyond those discussed in the lecture?
2. Does the inability to provide "complete" criteria for distinguishing the mental from the physical mean we really don't have a clear idea of what the mind is? If so, is that a problem?

Suggested Reading

Ludwig, Kirk. "The Mind-Body Problem: An Overview." *The Blackwell Guide to Philosophy of Mind.* Eds. Stephen P. Stich and Ted A. Warfield. Chapter 1, pp. 1–46. Oxford: Wiley-Blackwell Publishing, 2003.

Other Books of Interest

Dennett, Daniel C. Chapter 2: "Explaining Consciousness." *Consciousness Explained*. Boston, MA: Back Bay Books, 1992.

Descartes, René. "Meditations on First Philosophy (1641), Meditation 2." *Philosophy of Mind: Classical and Contemporary Readings*. Ed. David J. Chalmers. New York: Oxford University Press, USA, 2002.

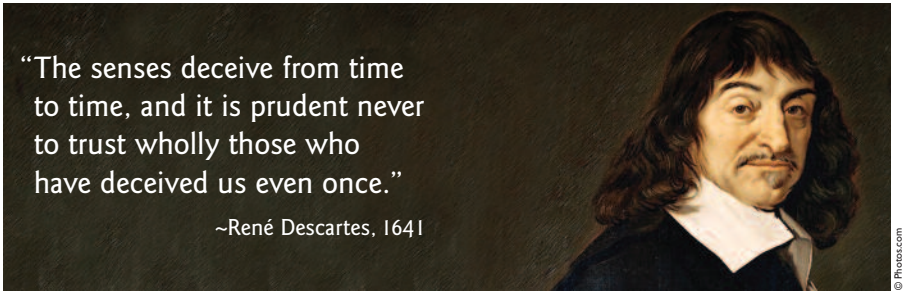
Websites of Interest

1. The *Scientific American* website provides an article entitled "The Mind-Body Problem: What Does It Feel Like to Have Too Many Arms?" by Moheb Costandi (April 14, 2009). — http://www.scientificamerican.com/article.cfm?id=body-integrity-identity-disorder
2. The BBC website provides a discussion of the mind-body problem from their *In Our Time* program (January 13, 2005). — https://www.bbc.co.uk/programmes/p003k9b8

Lecture 2

Dualism

The Suggested Reading for this lecture is René Descartes's "Meditations on First Philosophy (1641), Meditation 2" in *Philosophy of Mind: Classical and Contemporary Readings,* edited by David J. Chalmers.



Introduction

We now begin with our first major theory or doctrine concerning the relationship between mind and body.

Basic Positions: Dualism and Its Competitors

Dualism: the mind is completely separate from the body or brain, and is entirely non-physical in nature.

Physicalism (or *Materialism)*: ultimately only physical or material things exist.

Idealism: ultimately only mental things exist.

Idealism has a vigorous history in philosophy, but the major debate these days is between Dualists and Physicalists.

Two Varieties of Dualism

Substance Dualism: there are two basic types of things or objects in the world.

Property Dualism: there are two basic types of *properties* in the world.

Property Dualists hold that while mental properties are different than physical properties, the very same thing or object—say, the physical brain—can have both. One need not hold that in addition to the brain there also exists the "mind," as a substance or thing.

In contrast, Substance Dualism insists that mental properties are different from physical properties precisely because the things that have mental properties—minds—are very different from the things that have physical properties, such as bodies and brains. Minds and bodies are literally very different substances altogether.

Nowadays, most Dualists are Property Dualists, but the traditional form of Dualism, which still does sway many people today, is Substance Dualism. This is the position of the great French thinker René Descartes (1596–1650).

Cartesian Substance Dualism

The basic idea is simple: that minds and bodies, in particular brains, are totally different sorts of things or substances. What makes them "distinct" is that they can exist independently of each other, even if, during our lives, they exist simultaneously and "depend" on each other in various ways. Moreover, they are characterized by very different properties. According to Descartes, the defining property of "matter," or "body," is its spatiality: everything physical has properties such as size, shape, and a location. In contrast, minds have no spatial properties and are not even located anywhere in space; their defining property is simply "thought."

Roughly, a mind is that which is conscious and is subject to having qualia and intentional states.

Six Arguments for Substance Dualism

Argument from Religion

Any religious believer must take this argument very seriously. All the major Western religions accept that an individual may continue existing after the death of his or her body. But then the individual must be *more* than just his or her body. This additional thing is the mind.

Argument from Dreams and Afterimages

We began Lecture I by suggesting that afterimages were neither outside nor inside the physical head, and thus existed "in the mind." We can now generalize this to other phenomena, such as *dream image*s. Suppose, in a recent dream, you were enjoying a cool beverage on a hot sandy beach with the person of your choice. In reality, we may suppose, you were in your bed alone in the middle of a freezing winter night, with no sand in sight. Now ask yourself, just what was it you were perceiving in this dream? Not any physical objects, because none of those physical objects are there to be perceived. Nor would a brain surgeon find any such images in your brain. These dream images, too, must exist only in the mind—and the mind must be different from the brain.

Argument from Introspection

Think about something physical, such as your brain. You're pretty clear about what it's like to be "thinking of" or "aware of" this object. But now introspect a little—that is, think about your thoughts themselves, or about your *self*, the you who is doing your thinking. What we're

aware of in thinking about the brain is different from what we're aware of in thinking about our thoughts or our self—because the brain is itself different from the latter.

Argument from Defining Properties

According to Descartes, the defining property of a physical body is that it is spatial in nature, while the defining property of mind is "thought." But now if body and mind have different defining properties, they must be two different things.

Argument from Conceivability

Descartes writes:

> First, I know that everything which I clearly and distinctly understand is capable of being created by God so as to correspond exactly with my understanding of it. Hence the fact that I can clearly and distinctly understand one thing apart from another is enough to make me certain that the two things are distinct, since they are capable of being separated, at least by God . . . [Now] on the one hand I have a clear and distinct idea of myself, in so far as I am simply a thinking, non-[spatial] thing; and on the other hand I have a distinct idea of body, in so far as this is simply [a spatial], non-thinking thing. And accordingly, it is certain that I am really distinct from my body, and can exist without it ("6th Meditation," Chalmers, 2002, p. 16).

There's a lot going on here, but the basic idea is this: Whatever I can clearly and distinctly conceive of as being separate or separable is *in fact* capable of existing independently of each other. I can clearly and distinctly conceive of mind and body as being separable; therefore they are capable of existing independently of each other. Therefore they are two different things.

And on what basis, exactly, does Descartes think we can conceive of mind and body as being separable? Answer: our concepts of each of them, as we saw in argument four (defining properties), are very different concepts.

Argument from Inexplicability

Certain mental phenomena, it is claimed, cannot be adequately explained by purely physical science; therefore there are irreducible mental entities or properties. Descartes himself noted two such phenomena: the first was our mastery of language, the second was our ability to think generally, or use "reason." It follows, he believed, that there must be more to us than our purely physical bodies—we must also have minds. Nowadays, proponents of this argument focus on various aspects of consciousness.

Brief Reflection on the Arguments for Substance Dualism

These arguments come from a lot of different directions. This makes for a lot of ammunition for Substance Dualism and means that the position must be taken extremely seriously.

But while the arguments are all linked or related in various ways, they must each be considered individually. We'll now look at possible replies to each of the arguments above.

Some Replies to the Six Arguments for Substance Dualism

Argument from Religion

This argument only applies to those who are already believers in personal immortality and is thus somewhat limited in scope.

Argument from Dreams and Afterimages

This argument strikes many as compelling, but there may be ways of resisting it. Take a look at the real desk in front of you. No one doubts that what we see here is the physical thing, the "real desk." Now close your eyes and visualize the desk. In so doing you've created a memory image—and most are inclined to say that what we are visualizing *is* the real desk, the same physical desk we just looked at. In remembering something we are re-creating our original perception of it—but it is the same *it*, the same physical thing, we are re-perceiving. So having this experience doesn't mean "there exists some *mental* image"; rather, we are just re-perceiving the same physical thing we did previously.

But then why can't the same be true for afterimages and dream images? They are just new, different ways of perceiving previously perceived *physical* things.

Argument from Introspection

In reply, we might ask ourselves just how reliable introspection really is. Indeed, science often tells us that things are quite different from how they may seem to us. The moon looks larger when it's on the horizon, but it is not in fact larger: that's a perceptual illusion. So, too, it may "seem" to us that our selves and thoughts are different from our brains and bodies, but that intuition may be as false as the illusion that the moon is larger on the horizon.

Argument from Defining Properties

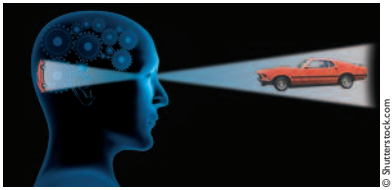
Is it really true that if X and Y have different defining properties then they must be different things? The defining properties of Superman might include that he is able to fly, or that he is super strong. Those of Clark Kent might be that he is a reporter and that he is dating Lois Lane. These surely are different defining properties—and yet Superman and Clark Kent, as we know, are one and the same person.

Two different ways of thinking does not therefore entail that we have two different things. And so too mind may be defined by different properties than body, but that may just mean we have two different ways of thinking about—two different concepts concerning—what is in fact one single thing.

Argument from Conceivability

The reply here is the same as to the previous argument. It's simply not true that "whatever we have two distinct concepts of are capable of existing

separately"—since the mind and brain might well be the same thing despite our having distinct concepts of them, just as in the Superman/Clark Kent case.



Argument from Inexplicability

Here, now, it cannot be denied that there are mental phenomena that currently cannot be explained by purely physical science: language, reason, and consciousness are surely far from being fully understood. But both neuroscience and cognitive science are quite young—and who can say, exactly, what these sciences will yield in the future? Perhaps they will, eventually, provide a complete physical explanation for all these mental phenomena. And so it would be premature to conclude that the mental is distinct from the physical.

Let's conclude now by looking at two arguments *against* Dualism, along with some replies.

Two Arguments Against Substance Dualism

Argument from Science

Contemporary cognitive science is very Physicalist in its theorizing; the theories that get developed aim precisely to explain how the physical brain generates mental phenomena. This science, despite being young, is already extremely successful. We thus have reason to believe that, in the long run, it will be able to explain everything mental in physical terms.

But the reply is simple: when physical science fully explains everything mental, then Physicalism shall rule the day. But right now it is merely a promissory note and, as we'll see in Lectures 11 and 12, contemporary Dualists offer reasons to believe this day will *never* arrive.

Argument from Causation

When neuroscientists look at the brain, it appears that everything there involves purely physical causes. If Dualists claim that the mind has causal effects on the brain, then that would make neuroscience impossible: one couldn't explain brain states in terms of earlier brain states, since additional factors would be contributing.

Further, causation between mind and brain seems literally unintelligible or inconceivable. If the mind is non-spatial in nature, how *could* it interact with the spatial brain?

Conclusion

In this lecture we've explored Substance Dualism. We first sketched the doctrine along with its primary competitor, Physicalism, and then looked at a half-dozen arguments supporting it, as well as some replies to those arguments; and then we looked at a couple of arguments against Dualism along with replies to *those* arguments.

FOR GREATER UNDERSTANDING

Questions

1. Which of the arguments for Substance Dualism do you find least and most convincing (if any)?
2. Which of the arguments against Substance Dualism do you find most (or least) convincing?

Suggested Reading

Descartes, René. "Meditations on First Philosophy (1641), Meditation 2." *Philosophy of Mind: Classical and Contemporary Readings*. Ed. David J. Chalmers. New York: Oxford University Press, USA, 2002.

Other Books of Interest

Goldberg, Sanford, and Andrew Pessin. Chapter 2: "The Mind-Body Problem." *Gray Matters: An Introduction to the Philosophy of Mind*. Armonk, NY: M.E. Sharpe, 1997.

Smythies, John R. *The Case for Dualism*. Ed. John Beloff. Charlottesville, VA: The University of Virginia Press, 1989.

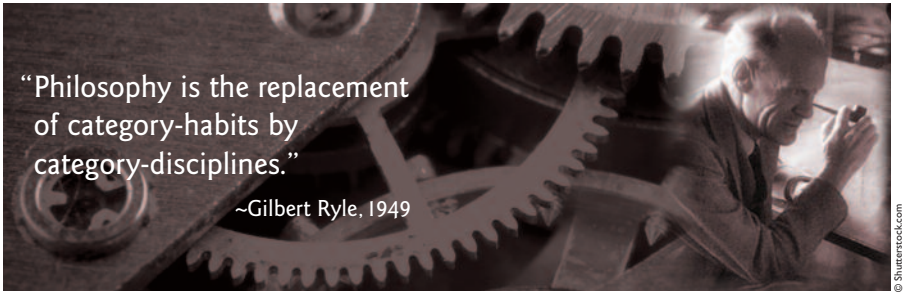
Websites of Interest

1. The *Internet Encyclopedia of Philosophy* provides an entry titled "René Descartes: The Mind-Body Distinction" by Justin Skirry (May 3, 2006). — http://www.iep.utm.edu/descmind
2. The *Stanford Encyclopedia of Philosophy* features a detailed entry on zombies as the concept applies to the mind-body problem (September 8, 2003). — http://plato.stanford.edu/entries/zombies

Lecture 3

Psychoanalysis, Behaviorism, and the Science of the Mind

The Suggested Reading for this lecture is Gilbert Ryle's *The Concept of Mind*, chapter 1, "Descartes' Myth."



Introduction

The early twentieth century saw the development of two diametrically opposed conceptions of the "science of the mind," that is, psychology.

Freud and Psychoanalysis

Sigmund Freud's Psychoanalysis asserts that the mind has an extraordinarily complex structure, with "regions" and "operations" such as the "ego," "id," "libido," "the unconscious," "repression," "sublimation," and so on. These terms make no reference to anything physical, such as the brain or specific brain operations, nor to physical behavior.

Thus Psychoanalysis may be a Dualism taken to its most sophisticated extreme: not only is the mind understood in terms completely distinct from anything physical or behavioral, but we now have an impressively complex theory of mental life attempting to explain nearly all aspects of human behavior.

As such, Psychoanalysis is a candidate for a Dualist "science of the mind."

But What Is Science, Anyway—and Is Psychoanalysis Really One?

In fact the question of whether Psychoanalysis is a "science" is hotly debated. Its defenders point to its ability to "explain" all sorts of human behaviors; interestingly, its critics point to the very same feature, arguing that Psychoanalysis shares that ability with such pursuits as astrology, which can also tell appealing "explanatory" stories—but typically only in retrospect. Like astrology, and unlike real science, it's claimed, Psychoanalysis is incapable of supporting reliable predictions, directly observable phenomena, or repeatable experiments.

It is this latter point that motivates the alternative conception of the science of the mind.

The "Science of the Mind"

The origin of psychology, as a *science*, is typically traced to the end of the nineteenth century. For it was only then that psychology turned from a purely philosophical "think about it" endeavor to one where "scientists" made experiments.

From early on, though, it was recognized that scientifically studying the "mind" wasn't such an easy thing to do.

And it surely isn't for a Dualist, according to whom the mind is private, internal, and accessible only by its owner; it's just not the kind of thing that more than one person could ever access or observe. It seems therefore to follow that no objective knowledge of its nature and operations is possible. And if not, there could be no such "science" of the mind.

What Dualism offers is "subjective" knowledge at best, a "privileged access" to our own internal thoughts, sensations, and feelings. You might think that this could support a kind of science of the mind, one based on "introspection." Except that even here there were problems. For while Dualists insisted that the mind has an infallibility and incorrigibility with respect to itself, it now began to be appreciated that introspection is surprisingly unreliable and difficult to achieve with any real precision. So not even subjective knowledge of the mind could support a science of the mind.

If there were to be a "science" of psychology, then, it must not consist in study of the "inner world" but in study of something about which there *could* be objective knowledge: namely, human behavior. And perhaps no thinker better expresses that orientation than B.F. Skinner.

B.F. Skinner and the Critique of Psychoanalysis

Skinner was, arguably, one of the two most famous psychologists of the twentieth century—along with his diametrically opposed counterpart, Sigmund Freud.

Skinner shared many of the complaints about Dualism we just mentioned. He was also deeply opposed to Freudian Psychoanalysis, for some of the general reasons we've sketched, including its invocation of unobservable mental phenomena intervening between the observable stimulus and behavioral responses. Skinner also objected to these for two particular reasons.

First, the unobservables of Psychoanalysis are always "underdetermined" by the observable evidence.

To illustrate this idea, consider the following scenario. You observe someone, sitting before some food, who reaches for the salt shaker and shakes it heartily upon his meal. Your task is to "explain" why that person just did that. Well, here's one "theory": he loves very salty food and believed the dispenser before him to contain salt, and thus shook it over his plate. This theory invokes two unobservable, inner, "mental" states: he "loves" something, and he "believes" something.

But now notice that many other different mental states might explain the very same behavior. Maybe he loves pepper, and believed the shaker to contain pepper. Maybe he likes ordinary amounts of salt, but believes the food to have been under-salted. Or maybe he believes he is being observed and simply wants to mess with your head by pouring ridiculous amounts of salt on his food. And so on.

The point is that any given bit of observed behavior can be explained by many different, wildly divergent, combinations of inner mental states. Thus, the observable facts "underdetermine" the mental theory. And if so, then we could never know just which is the "true" inner theory of the mind.

Second, even if we could somehow settle on the one single "correct" mental theory in a given case, the theory would be "explanatorily empty."

To illustrate, consider an oft-quoted example from the French playwright Molière (1622–1673)**:** in one famous scene he portrays a medical student being asked to explain how a certain potion puts people to sleep. The student responds by invoking the "soporific quality" of the potion—which sounds like a perfectly good explanation until you realize that "soporific quality" means "a quality tending to put people to sleep." If so, then the "explanation" amounts to saying that "its tendency to put people to sleep puts people to sleep"—which is no explanation at all. Invoking "soporific quality" is "explanatorily empty"—it doesn't explain what it's supposed to explain.

Well, Skinner suggests, purely "mental" theories such as Psychoanalysis amount to the same thing: because the mental phenomena are all unobservable, mental language can only be defined in terms of observable behavioral consequences. But then any mental explanation you offer of behavior will amount to merely invoking the person's tendency to behave in just that way, which would be no explanation at all.

Hence, Skinner concludes, a science of the mind needs something that leaves *out* all this inner mental talk, something that discusses only observable stimuli, conditioning, and observable behavior responses—in other words, Behaviorism.

Two Kinds of Behaviorism: Methodological vs. Philosophical

Methodological Behaviorism is a program for science, as we've just described—where scientists attempt to study only observable behavior and its various causes and conditions.

But *Philosophical* Behaviorism is a theory about what we "really mean" when we use various words for mental things, words such as "mind," "thought," "sensation," and so on. When we use such language, what we are *really doing* is making various sorts of comments about people's overall patterns of behavior. Our ordinary mental language only refers, really, to behavior—so we never even make reference to a separate private (Dualist) mind in the first place.

The most important Philosophical Behaviorist was Gilbert Ryle (1900–1976).

Gilbert Ryle and Philosophical Behaviorism

According to Ryle, the whole mind-body problem arises because we make a "category error": we treat mental terms as if they were in the same "logical category" as physical terms, when in fact they are not.

What does this mean, exactly? Well, ordinary physical terms refer to ordinary external things, located in the space surrounding us. By analogy, we think of mental terms as also referring to "inner" objects and states. By treating physical and mental language as being in the same "category," we treat the words to refer to exactly the same *sorts* of things—the only difference being that one is "external" and the other is "internal."

But now, Ryle argues, this is a mistake, for mental language belongs to another kind of logical category altogether from physical language. To illustrate this, he imagines an individual taking a tour of a university. After seeing all the buildings, the individual then asks, "Now, where is the university itself?"

Clearly there is something wrong with this question. As Ryle diagnoses the problem, the person simply does not know how to use the word "university." And something similar goes on when we treat mental language as if it referred to inner, non-physical objects. For just as "university" does not refer to some object distinct from the buildings constituting the university, so too mental words do not refer to things distinct from all the purely physical things that constitute reality.

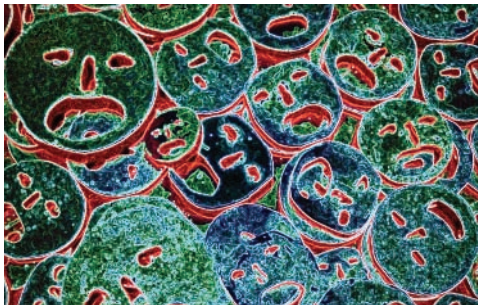
In fact, Ryle argues, our mental language is really an indirect way of referring to behavioral dispositions—so the goal of a philosopher is to give proper "translations" of mental terms into behavioral language. Once we do that, Ryle thinks, the whole mind-body problem evaporates—for we no longer have to explain how the two realms exist side-by-side, if, in fact, there aren't two such realms.

Some Examples of Philosophical Behaviorist Linguistic Analysis

When we say "Fred believes it's raining," it looks as if we're saying something about Fred's inner mental life—but really, the Behaviorist suggests, this sentence merely indicates such observable behaviors as that Fred is likely to carry an umbrella, wear a raincoat, or stay inside. Similarly, to say someone is angry is ultimately to describe their tendency to behave angrily; to say someone is virtuous is to say they tend to behave virtuously, and so on.

Two Points in Support of Philosophical Behaviorism

First, there does seem to be a close conceptual connection between mental states and behavior. For example, part of the very concept of pain, and thus the meaning of the word "pain," is avoidance-behavior: someone who didn't understand that pain was something to be avoided would simply not have the concept of pain as we understand it.



Second, it also seems that there could be no difference in mental states between two people without some sort of difference in their behavioral dispositions. Similarly, there could be no difference in *behaviors* between two people in the same circumstances without some difference in *mental states*. Mental states and behavioral dispositions are therefore very closely, intimately connected—just as the Philosophical Behaviorist insists.

Three Problems for Philosophical Behaviorism:

Philosophical Behaviorism is *very* counterintuitive.

Further, the translations are generally just unworkable, for several reasons:

First, it just doesn't seem possible to give truly satisfactory translations of mental sentences into behavior sentences.

Second, the translations seem unlikely to achieve the elimination of reference to mental terms altogether. For any satisfactory translation almost certainly will involve reference to various further mental terms.

Third, it seems easy to conceive of various counterexamples to the Behaviorist claim, cases where a mental term clearly does not refer to the corresponding behaviors.

Finally, we're aware of our own qualia and intentional states by introspection even when we're not aware of any particular behavioral dispositions. We understand these phenomena "internally," and the behavior-independent concepts so produced are what we have in mind in ascribing such states to others.

Ultimately these problems led to the demise of Behaviorism.

Conclusion

If there's to be a "science of the mind," it seemed, then it must be something objective and observable—and so Cartesian Dualism and Freudian Psychoanalysis won't count as scientific. Behaviorism, as a science of human behavior, fared much better, complemented by Philosophical Behaviorism. But in the end Behaviorism simply was not a successful theory.

FOR GREATER UNDERSTANDING

Questions

1. If Psychoanalysis is not a "science," per se, then what is it? Can it still be valuable, even "true," even if it's not strictly speaking scientific?
2. How might a Behaviorist respond to any of the "problems" for Behaviorism mentioned in the lecture?

Suggested Reading

Ryle, Gilbert. Chapter 1: "Descartes' Myth." *The Concept of Mind*. 60th anniv. ed. New York: Routledge, 2009 (1949).

Other Books of Interest

Goldberg, Sanford, and Andrew Pessin. Chapter 2: "The Mind-Body Problem." *Gray Matters: An Introduction to the Philosophy of Mind*. Armonk, NY: M.E. Sharpe, 1997.

Skinner, B.F. *Walden Two*. New rev. ed. Indianapolis, IN: Hackett Publishing Co., Inc., 2005 (1948).

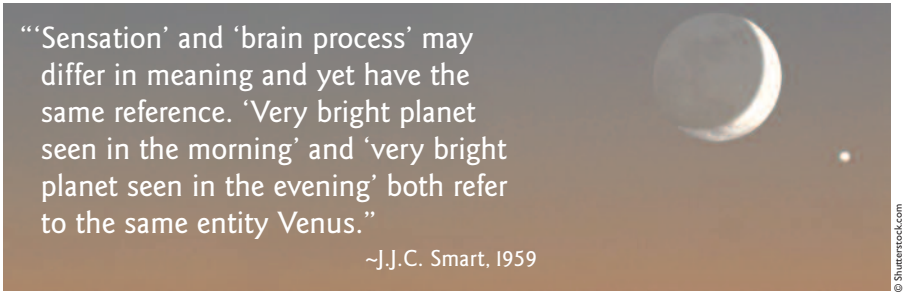
Websites of Interest

1. The B.F. Skinner Foundation was established in 1989 to publish significant literary and scientific works in the analysis of behavior and to educate professionals and the public about the science of behavior founded by Skinner. — http://www.bfskinner.org/BFSkinner/Home.html
2. The Hist-Analytic website by Stephen R. Bayne features the text of Gilbert Ryle's "Ordinary Language" as it appeared in issue LXII of *Philosophical Review* (1953). — http://www.hist-analytic.org/Ryle.htm

Lecture 4

The Identity Theory

The Suggested Reading for this lecture is J.J.C. Smart's "Sensations and Brain Processes" in *Philosophy of Mind: Classical and Contemporary Readings*, edited by David J. Chalmers, pp. 60–68.



Introduction

Though Behaviorism failed, Physicalism itself was not defeated. For one could claim that mental phenomena are just identical to other physical phenomena, in particular brain phenomena. We thus get the Identity Theory: "Mental states are identical to brain states."

There's no doubt that the Identity Theory is very intuitive to many. It seems to reconcile three important facts:

1. That we truly do have mental states.
2. That our mental states have *something* to do with our brains.
3. That our brains are purely physical entities.

Further, the Identity Theory fits nicely with our scientific knowledge of the biological development of human beings. What we observe is the development of a brain and brain activity; we *don't* observe any special point at which a distinct "mind" gets "added" to the developing being.

Finally, the Identity Theory fits very nicely with "Ockham's Razor," the methodological principle that "one should never multiply entities beyond necessity." One's goal is to explain as many different phenomena as possible with the simplest, most efficient theory. All else being equal, the Identity Theory clearly fits Ockham's Razor better than Dualism.

The Three Basic Assertions of the Identity Theory

The Identity Theory first asserts not merely that mental states and brain states are "correlated" with each other, but that mental states are *identical* to brain states. This means we will need some general theory about when, exactly, two apparently distinguishable things actually count as identical.

Second, the Identity Theory claims that all mental states are identical to brain states, *not* the reverse. There's no reason to think that every state of the brain gives rise to a mental state.

And lastly, the "kind" of identity between the mental and the physical is a "type" identity, not merely a "token" identity.

Types versus Tokens

We notice not merely that there are many different individual things existing in the world, but that these things often fall into categories. Here's a nickel, here's a dime, here's a dollar bill—all individual instances of the more general category of "money." Here's a bird, here's a rhinoceros, here's a fish—individual instances of the general category "animal." Then, too, there are categories of things that resemble each other, sometimes quite closely. Consider two electrons: these seem to be exactly identical to each other. Except of course that there are *two* of them.

What we say is that there are two "token" electrons, but only one "type" of thing: "electron." We may have two token dimes, but these are two tokens of the very same single type, "dime." And we may have some token nickels, dimes, and dollar bills, but these are all of the same single, more general, type, namely "money."

So "tokens" are individual instances of things, while "types" are the more general categories under which tokens may get classified.

Type-Identity versus Token-Identity

To assert that there's a "type-identity" in place is to assert that what appear to be two different "types" really are the same single type. Aristotle believed, for example, that the type "human being" is the same type as "rational animal." So, for him, there is just one type here called by the two different names.

But, now, if the type "human being" is identical to the type "rational animal," then every token (individual) human being must also be a token rational animal and vice versa. If two types are identical, in other words, all the tokens of the first type are also tokens of the second type and vice versa.

The Identity Theory as a Type-Identity Theory

Let's now apply this to the mind-brain case. The Identity Theory, as a Type-Identity Theory, is now claiming this: "Every type of mental state is identical to

some type of brain state." There are, of course, many "types" of mental states: there are different types of sensations, different types of intentional states, and indeed there are different sub-types of each of these.

Suppose Fred is feeling a very specific kind of pain. On the Identity Theory, his feeling that pain is identical with his brain being in a very specific state: let's call it N17 for short. When Fred is feeling some other kind of pain, he's in some other brain state: say, N22. When Fred is thinking that it would be nice if the Yankees lost, he is in some other type of brain state: say, N212.

And, of course, since this is a type-identity, the reverse holds as well: every time he feels that pain he is in N17, and every time he is in N17 he is feeling that pain.

The classic proponents of Identity Theory—philosophers U.T. Place and J.J.C. Smart—were Type-Identity Theorists. We'll focus on Smart's work; and since Smart proceeds by responding to objections to the Identity Theory, we'll follow suit.

Objection One to the Identity Theory

Mental terms are not synonymous with neural terms. So when I say that "I'm in pain," I am *not* saying "N17 is firing." So how could this pain be identical to N17, when in speaking of "pain" I'm speaking of something other than my brain?

Identity Theorist's Reply to Objection One

The mistake in this objection is that it overlooks an important fact about language.

In particular, we must distinguish between the *meaning* of a certain term and the object that the term refers to, its *denotation*. For example, the phrase "the Evening Star" denotes the very same object as "the Morning Star," namely, the planet Venus. But the phrase "the Evening Star" does not have the same *meaning* as "the Morning Star." After all, nobody holds that "morning" has the same meaning as "evening." So two phrases may differ in meaning even while sharing a denotation.

The Identity Theorist can now simply *grant* that the word "pain" is *not* synonymous with any neural expression such as "N17." All she is claiming is that the *denotation* of "pain" is identical to the *denotation* of "N17." And two terms may differ in meaning yet still share denotation.

Objection Two to the Identity Theory: The Leibniz's Law Family

These objections all invoke a principle about identity known as "Leibniz's Law," named for G.W. Leibniz (1646–1716). It states:

"If X is identical to Y, then whatever is true of X is also true of Y."

Or in the opposite direction,

"If something is true of X that is *not* true of Y, then X must not be identical to Y."

These objections all claim that there are things true of mental states that *aren't* true of brain states, or vice versa—and so claim, by Leibniz's Law, that the Identity Theory is false.

Objection Two-a

Mental states are "knowable by introspection," but brain states are not.

Objection Two-b:

An afterimage may be bright, and green, but of course no brain state is "bright" or colored in any way like that afterimage. So the after-image has properties that brain states don't—and thus cannot be identified with brain states.

Objection Two-c

Finally, brain states have physical properties that the corresponding mental states do not. For example, a molecular movement in the brain may be fast or slow, or circular or straight, but a given thought or perception is none of these. Or we might be dreaming of a lemon two feet in front of us, but the corresponding brain state isn't in front of us. Or our brain may be at ninety-eight degrees, but our belief that it's raining does not itself have a temperature.

Identity Theorist's Reply to Objections Two

According to Objection Two-a, mental states are "knowable by introspection" in a way that brain states are not. But the Identity Theorist's reply here is to flip the argument on its head: you only get that difference if you already *assume* that mental states are not identical to brain states. For if, in fact, they are, then what we are introspecting just *are* our brain states, even if we don't recognize them as such.

According to Objection Two-b, mental items such as afterimages have properties (such as colors) lacked by brain states. Here the Identity Theorist replies by observing that what she is claiming is not that afterimages *themselves* are identical to brain states, but rather the experience of *having* afterimages. And nobody is ever tempted to believe that the experience *itself* has any color.

According to Objection Two-c, finally, brain states have physical properties that mental states do not. But the Identity Theorist offers no objection to saying that our *experience* of an afterimage (say) is located (in one's brain), even if the afterimage itself experienced doesn't have a location. Similarly the activity of dreaming may well have a physical property such as a temperature even if the "things being dreamed" do not. Once we identify the brain states not with "what" is experienced but with the activity of experiencing itself, then we can grant the mental state of experiencing all the physical properties we like.

Putting these together, then, the Identity Theorist claims that Leibniz's Law does not preclude the identity between mental states and brain states.

Objection Three to Identity Theory

To say that something of a given type is "multiply realizable" is to say that a thing of that type might actually be realized in, instantiated in, or constituted by many different kinds of lower-level types. Tables are "multiply realizable," for you can make a table out of wood, of metal, of plastic, and so on. But then each of *these* can in turn be realized in different materials: different kinds of woods, different kinds of metal, and so on.

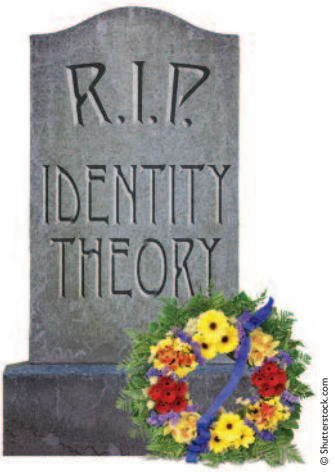
The higher-level concepts here are distinguishable from the particular form they might take on the lower level. What we mean by a "table" is something distinguishable from any particular material. And what that shows is that wherever something is multiply realizable, there is no type-identity between it and any of the things in which it can be realized. The type "table" cannot be identified with the type "wood" precisely because tables can be made of other materials.

The third objection to Identity Theory is then this: mental states are *also* multiply realizable. For the very same type of mental state can be manifest in very different sorts of brains. Not only human beings have mental states, for example, but so too do all sorts of lower animals. But since animal brains differ dramatically from ours, the mental states we share with such creatures could not be realized in the very same types of brain states. So we cannot accept a type-identity between mental states and our brain states.

Identity Theorist's Reply to Objection Three

According to the third and final objection, mental states are "multiply realizable." And here we offer a rather surprising response: namely, there isn't one.

This objection, in fact, more or less resulted in the demise of the Identity Theory: it just no longer seemed possible to identify a mental state type with any brain state type. Indeed the goal of philosophers of mind subsequently became the discovery of precisely what, say, tokens of the mental state of "believing that p" might have in common when they are realized or instantiated in brains of very different types.



Conclusion

In this lecture we examined the Type-Identity Theory. We explored a series of objections to the theory, based on the meanings of our words, on Leibniz's Law, and on "multiple realizability." We saw that the Identity Theorist had substantial responses to the first two, but that the third pretty much defeated the Identity Theorist.

FOR GREATER UNDERSTANDING

Questions

1. Which of the objections to Identity Theory strikes you as the most compelling? Alternatively, which of the replies offered to those objections strikes you as least compelling?
2. Does the fact that (say) an animal could experience the same sensation as a human being mean, in your view, that that sensation cannot be identified with a brain state or event in us?

Suggested Reading

Smart, J.J.C. "Sensations and Brain Processes." 1959. *Philosophy of Mind: Classical and Contemporary Readings*. Ed. David J. Chalmers. Pp. 60–68. New York: Oxford University Press, USA, 2002.

Other Books of Interest

Goldberg, Sanford, and Andrew Pessin. Chapter 2: "The Mind-Body Problem." *Gray Matters: An Introduction to the Philosophy of Mind*. Armonk, NY: M.E. Sharpe, 1997.

Place, Ullin T. "Is Consciousness a Brain Process?" 1956. *Philosophy of Mind: Classical and Contemporary Readings*. Ed. David J. Chalmers. Pp. 55–60. New York: Oxford University Press, USA, 2002.

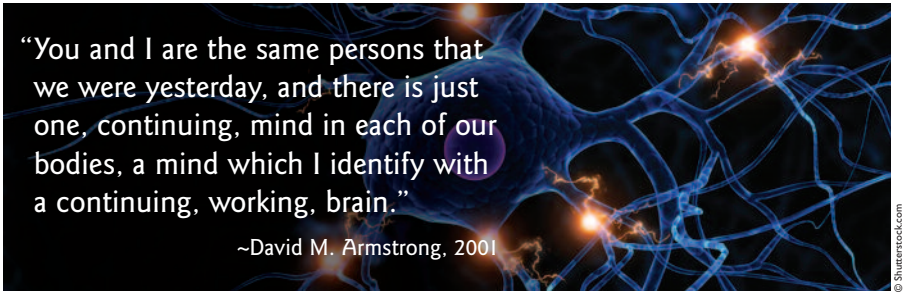
Websites of Interest

The School of Philosophy at the Australian National University provides information on the annual "Jack Smart Lecture" that has featured notable philosophers from around the world. — http://philrsss.anu.edu.au/smart-lecture.php3

Lecture 5

Functionalism

The Suggested Reading for this lecture is David M. Armstrong's "The Causal Theory of Mind" in *Philosophy of Mind: Classical and Contemporary Readings*, edited by David J. Chalmers, pp. 80–87.



Introduction

The last lecture introduced the idea of multiple realizability—that the very same type of mental state might be realized in many different sorts of underlying types of brain states. This point sets the stage for our next theory, called Functionalism. We begin with the basic concept of a "functional state, or role."

Two Examples of "Functional Roles"

The Concept of a "Poison"

What makes something a "poison" is that it has fairly well-specified causal consequences: when a creature ingests it, it causes various undesirable effects. So understood, many different substances can count as poisons. The concept of a poison is defined, in short, by a certain causal or "functional role" a substance plays. And we can easily distinguish the "functional role" so defined from the individual thing that may fill or play that role: the "functional role" is that of "having a tendency to cause discomfort, illness, or death in species," while the individual thing occupying that role might be a chemical, a bacterium, or a venom.

The Concept of "President of the United States"

This concept, too, is defined "functionally," that is, in terms of very well-specified procedures. *The* "United States President" at a given time is that individual who occupies that well-specified functional role.

Mental States as Functional States

We can now state the basic idea of Functionalism. It's this: that mental states are functional states. And to say this is to say that mental states are to be defined in terms of their functional roles.

And to say *this* is to say that what defines a particular mental state type is the causal relations that that state type bears to bodily or sensory input, other types of mental states, and ultimately behavioral output. What makes a given state a particular type of qualia state, such as "pain," is that it typically is caused to result from various sorts of bodily injuries; it typically causes other types of mental states (such as misery and unhappiness) in turn, and it typically causes various specific behaviors to ensue. Similar considerations apply to intentional states.

Functionalism need place absolutely no constraints on what might occupy these functional roles—what it is that "realizes" or "instantiates" the mental state in question—so that *any* system can "be in pain" or "believe that it's raining," as long as the system can have states that fit the relevant causal descriptions.

Functionalism, Behaviorism, and Identity Theory

Like Behaviorism, Functionalism understands that mental states are closely (or "conceptually") related to bodily input and our behavioral output. But unlike Behaviorism, Functionalism is quite willing to include other mental states in the relevant definitions. So unlike Behaviorism, which tried (and failed) to eliminate mental states, Functionalism isn't trying to eliminate the mental—and thus isn't failing.

Yet like both Behaviorism and Identity Theory, Functionalism is interested in maintaining Physicalism. How does it do that? While the functional role (that is, the definition or concept of the mental state) does not eliminate the mental, the theory can be Physicalist by suggesting that whatever it is that actually fits that definition or fills that role *is* physical.

But the key advantage Functionalism enjoys over the Identity Theory is that it gives up the problematic "type-identity" between the mental state type and the brain state type. No one demands a type-identity between the type "poison" and whatever it is that *is* the poison; the type "poison" just doesn't map, one-to-one, with (say) the type "chemical structure C," since things other than chemical structure C can be poisons. So, too, there need be no type identity between a mental state type and anything physical such as a brain state type. "Pain" might be "multiply realized," in other words.

Two Advantages of Functionalism

Functionalism first avoids the limitations of Behaviorism and Identity Theory—the latter, in particular, by accommodating multiple realizability.

A functionalist psychology can be autonomous with respect to lower-level sciences. Psychology can proceed, in other words, largely independently from, say, neuroscience. Why is that? Because if mental states are functional states, then it doesn't matter what the underlying substance is that realizes or instantiates those states.

Indeed, this same autonomy underlies the whole idea of Artificial Intelligence (that is, the idea that computers might be programmable so as to count as having genuine mentality). Functionalism says that the computer doesn't *have* to be just like the brain physically in order to instantiate mental states: it merely has to have states occupying the same causal roles as our own mental states do, and that is the job of programmers to bring about.

Problems for Functionalism

Broadly speaking, the problems fall into two categories: those dealing with "intentional states" and those dealing with "qualia." We'll briefly present the first category and then preserve detailed discussion until our next lecture; we'll focus here on the problems concerning qualia.

There are basically two strategies for objecting to Functionalism. Functionalism claims that mental states are functional states. That claim suggests that (1) it's impossible for something to be in a given mental state if it lacks something with the appropriate functional role and (2) it's impossible *not* to have the relevant mental state if a thing *is* in a state with the appropriate functional role. The two strategies for objecting to Functionalism correspond to challenging each of these two suggestions.

Functionalism and Intentionality

Functionalism, it is objected, cannot give an adequate account of intentional states. And the main strategy is the second one just mentioned: it seems possible to conceive of cases where something has states with the appropriate causes and effects, but lacks the relevant mental states.

The most famous version of this objection was stated by Berkeley philosopher John Searle. Searle introduced a thought experiment in which, he claimed, there was clearly a system that would count as "functionally equivalent" to a person who genuinely understood the Chinese language. To say that is to say that it exhibited states that had all the appropriate causes and effects of the states of a fluent speaker of Chinese. But, Searle argued, the system merely *mimicked* genuine "understanding" of Chinese—it did not in fact display the true mental state of understanding.

And so we have, in Searle's example, a strong candidate for a case where a system has all the right functional states, yet lacks the corresponding mental state. If so, then mental states cannot be identified with functional states.

Functionalism and Qualia

The charge against Functionalism here is that it cannot accommodate qualia. This charge can be made by way of either of the two strategies mentioned.

Consider some feeling of excruciating pain. To be sure, feelings of pain are normally part of an easily specifiable causal network. But it seems easy to sever states like pains—qualia in general—from their typical causes and effects, at least conceptually. Imagine stepping on a nail and feeling that excruciating pain in your foot. Now focus on the pain itself and conceive of experiencing it even without having stepped on the nail. But then the *concept* of pain itself does not intrinsically involve any connections to the normal causes of pains, contra Functionalism.

But if "pain" is not, after all, defined by its typical causes and effects, what is it defined by? Simply: by how it feels, what it's like to be in pain. But then Functionalism simply fails to give the proper account of qualia.

Is it possible, now, to imagine states having all the expected causes and effects yet not being the mental state in question—or even somehow failing to be genuine mental states with genuine qualia altogether?

Inverted Qualia

You and I both look at some grass and say "green," and look at the sky and say "blue." But how do we know that our inner sensations in fact are the same, or even similar? Perhaps what you see when you look at grass is in fact yellow—except that you have learned to call things looking that color "green." Similarly, what you see when you look at the sky looks green—except that you have learned to call things looking like that "blue." In this scenario, where our qualia are inverted relative to each other, we might still behave in all the same ways and yet what we're experiencing couldn't be more different.

As long as this scenario is *conceivable*, then Functionalism is in trouble: because, according to Functionalism, if two states have exactly the same "functional role," then they would be the very same type of mental state. And yet here we can imagine two states with the same functional role being about as different as two mental states could be.

Absent Qualia

A "zombie," to a philosopher, is a being whose behavior might be very much like our own but who is, in fact, mindless. And the conceivability of such beings raises problems for Functionalism.

The most colorful treatment of this issue is from New York University philosopher Ned Block, who imagines the following scenario:

> Suppose we convert the government of China to [F]unctionalism, and we convince its officials that it would enormously enhance their international prestige to realize a human mind for an hour. We provide each of the billion [plus] people in China . . . with a specially designed two-way radio that connects them in the appropriate way to other persons and to [some] artificial body . . . [fitted] with a radio transmitter and receiver . . . Surely such a system is not physically impossible. It could be functionally equivalent to you for a short time, say an hour (Chalmers, 2002, p. 96).

Block's idea, in short, is that the entire population of China might functionally mirror the neurons in your brain. But there is an enormous difference between you (or your brain) and the population of China as a whole: you (or your brain) is conscious, aware, has mental states, but surely the population of China, taken as a whole, does not.

But if so, then here we have a system with states displaying all the same "causes and effects" as the states of a human brain, yet without the corresponding mental states, the qualia. Qualia, again, are not defined by their causal roles, but by how they "feel." But then Functionalism fails to explain qualia.

Responses to These Problems?

There is, of course, more to be said here. In the next lecture we'll explore the responses to Searle's critique of Functionalism's account of intentionality in some detail. And in Lecture 14, we'll explore a position known as "Eliminativism"—which seeks to deny the reality of the mental altogether. We'll see there some ways to resist the qualia-based objections to Functionalism we've just explored. So the debate shall carry on.

Conclusion

In this lecture we sketched the basic idea of "Functionalism," according to which mental states are functional states and thus multiply realizable in different physical substrates. But we saw that Functionalism has problems accommodating both intentional states and qualia. Whether the former can be addressed is the topic of our next lecture.

FOR GREATER UNDERSTANDING

Questions

1. Is it really true that creatures with very different types of brains can be in *exactly* the same type of mental states? Or ought we to reject the possibility of the "multiple realizability" of mental states altogether?
2. How plausible are the colorful thought experiments in this lecture—the possibilities of "inverted qualia," and "zombies," and the nation of China's "collective mentality"? If they are implausible, does that undermine the objections to Functionalism?

Suggested Reading

Armstrong, David M. "The Causal Theory of Mind." 1981. *Philosophy of Mind: Classical and Contemporary Readings*. Ed. David J. Chalmers. Pp. 80–87. New York: Oxford University Press, USA, 2002.

Other Books of Interest

Block, Ned. "Troubles with Functionalism." *Philosophy of Mind: Classical and Contemporary Readings*. Ed. David J. Chalmers. Pp. 94–98. New York: Oxford University Press, USA, 2002.

Goldberg, Sanford, and Andrew Pessin. Chapter 2: "The Mind-Body Problem." *Gray Matters: An Introduction to the Philosophy of Mind*. Armonk, NY: M.E. Sharpe, 1997.

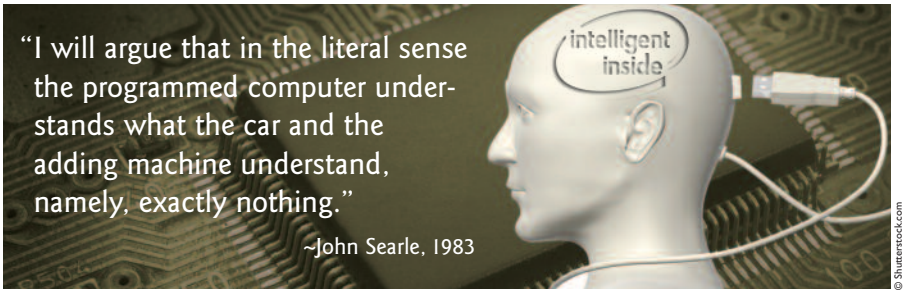
Websites of Interest

1. The *Saxelab* is an ongoing research project at the Massachusetts Institute of Technology that specializes in studying the human brain as it relates to Theory of Mind. — http://saxelab.mit.edu/index.php
2. Ian Thompson (Physics Department, University of Surrey, Guildford, UK) provides a paper entitled "Dualism: A Causal Correspondence Theory," based on the work of David M. Armstrong. — http://www.newdualism.org/papers/I.Thompson/qmc1h.html

Lecture 6

Computer Minds

The Suggested Reading for this lecture is John Searle's "Can Computers Think?" in *Philosophy of Mind: Classical and Contemporary Readings*, edited by David J. Chalmers, pp. 669–675.



Introduction

In the last lecture we saw that Functionalism, in allowing the possibility of the "multiple realizability" of mental states in different physical substrates, lays the groundwork for the possibility that computers might have genuine mental states. In this lecture we shall explore that possibility in some detail.

The Surprisingly Long History of Artificial Intelligence

The great ancient Greek philosopher Aristotle (384–322 BCE) proposed that logic—the backbone of rational thinking—can be "formalized." That is to say that good and bad arguments come in well-defined patterns that can be represented by symbols. For example, one general form of valid argument might be represented this way:

All Xs are Ys
All Ws are Xs
Therefore: All Ws are Ys

It doesn't matter what the particular argument is about; all that matters is the "form" of the argument. Aristotle's key idea was that thinking might ultimately be a matter merely of moving symbols around in various forms or patterns.

The More Recent, But Still Pretty Distant, History of Artificial Intelligence

The next major steps in the project of Artificial Intelligence consisted in attempts to build purely mechanical devices capable of doing mathematical calculations. John Napier (1550–1617) invented a set of ivory sticks to calculate

additions and multiplications. Blaise Pascal (1623–1662) designed a mechanical calculator that, it was said, could do the work of six accountants. G.W. Leibniz (1646–1716) invented a mechanical device capable of doing addition, subtraction, multiplication, and division. Charles Babbage (1791–1871) invented machines that could perform calculations out to thirty-one digits, including one that could actually be programmed to do different calculations by inserting punch cards. Finally, Bertrand Russell (1872–1970) and Alfred North Whitehead (1861–1947) argued that all mathematics was reducible to some basic bits of logic. If so, then all mathematical calculations and theorems could be provable by purely mechanical processes—by "computation." And if the essence of thinking is expressed by logic, then all thinking too could be explained in purely mechanical terms—as the manipulation of symbols according to well-defined rules.

Thus the idea was born of the mind as being a kind of computer—and of the possibility of a computer being programmed so as to have a mind.

Alan Turing and the Foundations of Artificial Intelligence

The British mathematician and logician Alan Turing (1912–1954) is considered the father of the modern computer. In a landmark 1950 article titled "Computing Machinery and Intelligence," he explicitly addressed the question, "Can machines think?" He offered an explicit criterion for determining whether a machine—a computer—may be said to have a mind: can the computer pass what is now known as "The Turing Test"?

The Turing Test

To put the idea simply: the question is whether a computer could ever be programmed in such a way that its behavior could fool a human interrogator into thinking that the computer was, in fact, a person, and so to "pass the Turing Test." The basic idea is that a computer should be recognized as having genuine mental states when its behavior is indistinguishable from that of a human being.

Since 1990 there has been an annual Turing Test competition—the Loebner Prize competition—in which programmers submit programs designed to fool judges. But no program has yet been judged to pass the Turing Test.

Some Objections to the Turing Test, and Replies

Turing himself considered a number of objections to his prediction that, in time, computers would pass the Turing Test and thus count as having minds.

In particular, there are various objections based on different disabilities: that is, claims that no computer would be able to do X, where X might include such things as being kind, resourceful, friendly, having initiative, falling in love, thinking about oneself, doing something really original, and so forth. But most of these denials, Turing notes, are asserted without any real argument. Further, most of them were made when computer technology was in its infancy; that's hardly an argument that no machine will *ever* display them.

Then there is what Turing calls the "Argument from Consciousness." The underlying idea is that computers merely process symbols; and while they might one day *imitate* various human behaviors and so pass the Turing Test, the computer itself wouldn't actually possess the mental states with which we produce our behavior.

In response, Turing observes that the most extreme form of such an objection would be the view that "the only way by which one could be sure that [a computer] thinks is to be the [computer] and to feel oneself thinking." And of course, that's an unanswerable objection, for no one is capable of doing that. But it's also clear that this objection is simply too extreme, for it would apply equally to our judgments about whether other human beings have minds.

But in fact we take others' behavior as sufficient evidence for their possessing minds. It would be unreasonable to demand any more with respect to computer mentality. The Turing Test is a legitimate test of mentality—for it is the one we use every day, in ascribing minds to other persons.

From Turing to Searle

In 1980, Berkeley philosopher John Searle published his notorious article "Minds, Brains, and Programs." Its heart was a thought experiment in which Searle imagined a system whose behavior was indistinguishable from that of a human being. But, Searle argued, it could be shown that this system merely mimicked genuine mentality but did not truly possess it.

"Weak A.I." versus "Strong A.I."

"Weak A.I." is the idea that computers can be useful tools in helping us understand the mind. Searle has no objections to this idea. But "Strong A.I." is the idea that computers can, when properly programmed, count as genuinely having minds or mental properties. And it is this idea that Searle rejects.

Crucial to Strong A.I. is an understanding of just what constitutes something as being a "computer." Searle notes that key to the concept of a computer is "that its operations can be specified purely formally," in terms of symbols being manipulated according to rules. Such symbols are specified without any reference to their semantics, to "what they mean." So a typical computer program might specify that certain patterns of electricity are to be replaced by other patterns of electricity, with no reference to what those patterns represent.

But now, Searle argues, this fact is fatal to the idea that mental processes are just types of programs. For genuine mental processes (such as our intentional states) are meaningful, have meanings, are "about" things. Mentality crucially involves semantics—the very thing that computer programs lack. So no computers could have genuine minds.

A famous thought experiment shows why.

The Chinese Room

Imagine that some computer programmers have written a program to simulate the understanding of Chinese. When one types in a question in Chinese, it produces appropriate answers to the question, in Chinese—and it passes the Turing Test, fooling native Chinese speakers into believing they're having a conversation with a genuine person.

But next imagine you are locked in a room with some baskets filled with Chinese symbols—symbols you don't understand. Through a slot in the door come slips of paper with Chinese symbols on them. You have a little rule book, in English, which tells you what to do with them. A rule might say: "When a squiggle comes in, take a squoggle and pass it back out." Nothing tells you what these symbols mean; you're just following these meaningless rules.

But next imagine that the symbols being passed into the room are called "questions" and the symbols you're passing back out are called "answers." And imagine further that the answers you've been sending out are indistinguishable from those of a native Chinese speaker. Imagine, in other words, that your activities in this room amount to passing the Turing Test.

All that, and yet you don't understand a word of Chinese.

The conclusion is clear. No computer can have a genuine mind—for all a computer does is precisely what you do in this room, namely process meaningless symbols, while genuine minds have genuine understanding.

Two Objections to the Chinese Room Thought Experiment

Objections to Searle came from every direction, from every discipline. But as Searle put it, they're all inadequate because his argument rested on a very simple truth: formal symbols are not sufficient for meanings, and computer programs only deal with formal symbols.

The Systems Objection

Searle's thought experiment wrongly focuses all our attention on the person in the room, who clearly doesn't genuinely understand Chinese. But there's more to the computer than just "that person": rather, there is a whole system in place, a system consisting of that person but also of the symbols, and of course the rule book containing the "program" specifying which symbols should be output. And it's *the system as a whole,* including the rule book, that understands Chinese.

The Robot Objection

Searle's thought experiment does show that a certain kind of system doesn't genuinely understand Chinese—but this isn't one that anyone would think could genuinely understand Chinese anyway. Rather, a more realistic system would include the ability to move around in the world, perform various tasks, and so on. In short, we should consider a computer program that operates a robot—a device that could successfully navigate the world as a human being does. That system would, if passing the Turing Test for its behavior, count as having a genuine mind.

Searle's Responses

The Systems Objection

Searle's response is simple. Imagine that you are in the room and have memorized the rule book and all the symbols. You now *are* the system, you have internalized it, everything in the system is now in you. What you know how to do is send out certain squiggles when certain squoggles come in. You pass the Turing Test—your responses are like those of a native speaker. Except that you *still* don't understand Chinese.

And why not? Because you're still just dealing with formal symbols. What's lacking is the meaning, the semantics, essential to true mentality. So even the "system as a whole" doesn't understand Chinese.

The Robot Objection

Searle's response here is similar to his earlier response. Imagine you're back in the Chinese room, manipulating those meaningless squiggles. But now unbeknownst to you, some of those incoming symbols are generated by television cameras attached to a robot's "eyes," and some of your outgoing symbols move robotic arms and legs in various ways. Imagine, in other words, that the Chinese room simply *is* navigating a robot through the world. Yet you *still* don't understand a word of Chinese. So even a system hooked up to a robot wouldn't have genuine mentality.

Conclusion

In this lecture we've looked at the project of Artificial Intelligence, at the claim that properly programmed computers could count as genuinely having minds. We sketched the long history of this project, and focused on the "Turing Test," which Searle's Chinese Room thought experiment attacks at its core. Even if a computer's behavior were indistinguishable from a person's, Searle argues, it still wouldn't genuinely have a mind. What seems essential to mentality, he insists, is the ability to grasp meanings, and that's just what computers can't do.

We'll pick up Searle's very contemporary idea—that minds are deeply linked with *meanings*—in Lecture 8.

FOR GREATER UNDERSTANDING

Questions

1. What sort of evidence could convince you that a programmed computer genuinely has a mind, other than its "behavior"? Is there any reason to distinguish between the criteria we use to conclude that other people have minds and those we might use for computers?
2. Searle's argument seems to suggest that there is more to genuinely understanding a language than simply knowing what output to give in response to various inputs. Is that correct? What does "understanding" consist in, if *not* just that?

Suggested Reading

Searle, John. "Can Computers Think?" 1983. *Philosophy of Mind: Classical and Contemporary Readings*. Ed. David J. Chalmers. Pp. 669–675. New York: Oxford University Press, USA, 2002.

Other Books of Interest

Goldberg, Sanford, and Andrew Pessin. Chapter 2: "The Mind-Body Problem." *Gray Matters: An Introduction to the Philosophy of Mind*. Armonk, NY: M.E. Sharpe, 1997.

Articles of Interest

Turing, Alan M. "Computing Machinery and Intelligence." *Mind*, 59, 1950, pp. 43–460.

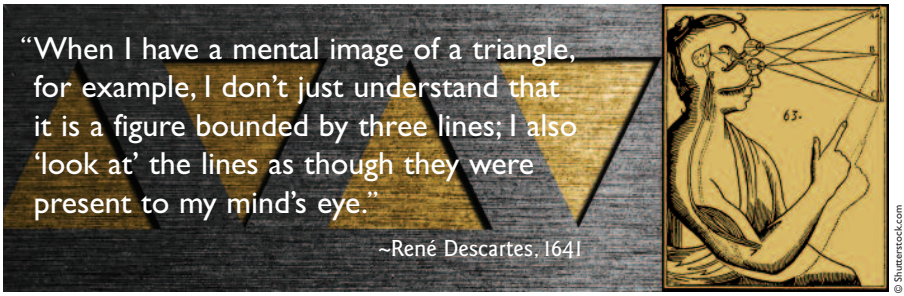
Websites of Interest

1. The *Loebner Prize in Artificial Intelligence* website provides information on the annually awarded Loebner Prize and a link to the full text of Alan Turing's article "Computing Machinery and Intelligence." — www.loebner.net/Prizef/loebner-prize.html
2. The *Machines Like Us* website features an interview with John Searle from 2009 in which he discusses the Chinese Room Experiment. — http://machineslikeus.com/interviews/machines-us-interviews-john-searle
3. The University of California Television website provides an episode of its interview program "Conversations with History: John R. Searle" from September 1999 on YouTube. — http://www.youtube.com/watch?v=giwXG3QYWQA

Lecture 7

Mind-Body Causation, Part I

The Suggested Reading for this lecture is René Descartes's "Meditations on First Philosophy (1641), Meditation 6" and "Passions of the Soul (1649)" in *Philosophy of Mind: Classical and Contemporary Readings,* edited by David J. Chalmers.



Introduction

No aspect of the relationship between the physical and the mental is more obvious, more important, or more problematic than the *causal* relationship between them.

Mind-Body Causation Is "Obvious"

Mind-Body

Mental states or events seem to cause physical states or events. For example, I find myself feeling hungry, a mental sensation, and then my body begins moving towards the fridge.

Body-Mind

Physical states or events seem to cause mental states or events. For example, you're listening to this lecture. Air molecules are moving, and this vibrates your ears, which makes your neurons fire—but then you have the mental sensation of hearing my voice.

Intra-Mind

Mental states or events cause various *other* mental states or events. For example, you might believe one thing "because" you believe another, or a thought might cause a desire.

Mind-Body Causation Is "Important"

Why? Because it seems essential to our most basic conception of what we are, as human beings, distinct from all other beings.

Inanimate objects are essentially passive, but we are capable of being "agents" who initiate actions. Further, unlike many things, we are capable of having genuine perceptions of the world. And finally, unlike perhaps all other things, we are largely *rational* beings.

Mind-body causation is in turn essential to each of these.

To be an "agent," to "initiate actions," is for us to act intentionally; and to act intentionally is for our mental beliefs and desires to cause our physical behavior.

To genuinely perceive something is for that physical thing to cause the mental perception.

To be rational is for one's thoughts to "hang together" in a certain way; and for that to occur requires thoughts to cause each other by "intra-mind" causation.

Mind-Body Causation Is Problematic

That is the general theme of this lecture.

The Mind-Body Problem

No problem for Dualism has had more historical traction than the causation problem, which is generally just called *the* mind-body problem. If your mind really is completely distinct from your body and brain, then how, exactly, can or do they causally interact? How could mental states or events cause physical states or events, and vice versa—especially if the mental is non-spatial in character?

Descartes's Misunderstood Response to the Mind-Body (Causation) Problem

The way the story is often told, Descartes attempted to solve the mind-body problem by asserting that mind-body causation occurred "at" the pineal gland in the brain. But of course that doesn't *solve* the causation problem. We still have no explanation for how a physical state of the gland can cause a mental state, or vice versa.

The real story is more subtle than this. But first let's explore two alternative views on mind-body causation from Descartes's own era.

Alternative Causal Theory #1: Nicolas Malebranche and "Occasionalism"

Nicolas Malebranche (1638–1715) was a French rationalist philosopher known best for his doctrine of "Occasionalism," which asserts that no ordinary things—like minds and bodies—have genuine causal powers. Instead, it holds, God alone is the single true cause of everything that occurs. So your desire for a snack results in your body's moving towards the fridge, but your desire doesn't cause your body to move. Rather, your desire is merely the "occasion" upon which God *Himself* causes your body to move.

Occasionalism dissolves the mind-body problem. There's no need to explain how mind and body interact, because they don't.

Occasionalism is also not as ad hoc as it may sound. Malebranche offers numerous arguments that make it more broadly applicable than merely to the mind-body case. Rather, it's a general theory of causation: *no* ordinary objects like minds and bodies have true causal powers—so no purely physical causation is possible either.

Alternative Causal Theory #2: G.W. Leibniz and "Pre-Established Harmony"

Imagine two grandfather clocks, exactly identical, wound the same way, and set to the same time. What we'd expect to see is a perfect correlation between the two clocks: when the first reads (say) 12:13, so too would the second, and so on. But this perfect correlation is not due to any causal relationship between the clocks. Rather, their respective readings are caused purely internally, by the internal mechanisms of each clock. The correlations are obtained because they have been externally coordinated. They run in perfect parallel, yet are causally separated from each other.

That is Leibniz's picture of the world as a whole.

There *appears* to be plenty of causal interaction between minds and bodies. But this appearance is an illusion: your desire for a snack may be correlated with your body's motion towards the fridge, but does not in fact cause it. Instead, each individual substance causes its *own* sequence of internal states. Your mind causes itself to go through a sequence of mental states, including desires, thoughts, perceptions, and so on. It is running on a purely internal program, much as each grandfather clock did in the preceding example. Meanwhile the physical world, including your body, is running on *its* own internal program, going through a sequence of states. And those sequences of events are all perfectly correlated: at the moment your mind's program produces its desire for food, your body's program produces the motion towards the fridge.

This "Pre-Established Harmony" also dissolves the mind-body problem by denying mind-body causation altogether.

Leibniz believes, not surprisingly, that such a system could only have been set up by God. But while the doctrine might appeal most to theists, it doesn't require belief in God. In fact, Leibniz offers many non-theist arguments for it.

With these two alternatives on the table, let's return to the question of what, precisely, Descartes's own actual theory is.

Descartes's *Real* Position on the Pineal Gland

Descartes invoked the pineal gland as the *location* for the mind-body problem. Why?

There's a certain problem in understanding perception. We have two eyes but we experience not two distinct images but rather one unified visual perception. The problem is how and why the multiplicity of sensory organs here

does not result in a multiplicity of perceptual experiences. More generally, we have multiple senses, and yet we are very good at realizing which sensory inputs "belong together": we easily recognize that the garlic clove we're seeing and the one we're touching and smelling are all the same. But how, if the visual and tactile and olfactory information are all traveling along distinct pathways?

One natural theory was that perhaps the information from the two eyes gets merged somewhere in the brain, and then in turn the information from all the senses gets merged somewhere as well.

According to the understanding of brain anatomy in the seventeenth century, much of the brain was symmetric, with many of its parts duplicated: today, for example, we speak of the left and right hemispheres that look to be roughly duplicates of each other (even though we now know they differ in many ways). The pineal gland stood out as a brain organ that was *not* duplicated and yet was closely connected to both hemispheres. Thus Descartes reasoned that the pineal gland might be where distinct information from all the different sensory organs came together. And if so, wouldn't that be the most logical location for the physical brain to convey that information to the mind—in order to generate the single unified perception we typically enjoy of the world?

For Descartes, the pineal gland was the "where" of mind-body interaction. But he was not attempting to explain the "how" of mind-body interaction.

The Plot (and the Problem) Thickens

There's a surprising passage in the final meditation in Descartes's most famous work, *Meditations on First Philosophy*:

> Nature also teaches me, by these sensations of pain, hunger, thirst and so on, that I am not merely present in my body as a sailor is present in a ship, but that I am very closely joined and, as it were, intermingled with it, so that I and the body form a unit. If this were not so, I, who am nothing but a thinking thing, would not feel pain when the body was hurt, but would perceive the damage purely by the intellect, just as a sailor perceives by sight if anything in his ship is broken. Similarly, when the body needed food or drink, I should have an explicit understanding of the fact, instead of having confused sensations of hunger and thirst ("6th Meditation," Chalmers, 2002, p. 17).

Dualism to this point seemed very much to be captured by the analogy of a sailor in a ship: the mind observes what is going on in the body, and in general controls or directs the body. And yet here Descartes rejects that analogy, with some interesting observations. If the mind were truly and completely distinct from the body, then we might perhaps "know" when our body suffered an injury; but we do more than merely "know"; we *feel* the pain. Similarly instead of merely knowing that our body needs food, we *feel hungry*.

This led Descartes to believe that mind and body are not completely separate after all—despite the Dualism. Rather, he said, they are "closely joined" or "intermingled."

What Descartes was invoking here are "qualia." If we want to appreciate Descartes's *real* theory of mind-body causation, we must recognize that he doesn't quite take Dualism to mean that the mind and body were utterly separate from each other. Rather, the existence of qualia is going to play some important role.

Descartes's *Real* Position on the Mind-Body Problem

Descartes's real position on mind-body causation was this: we don't know how mind and body interact. Or perhaps more accurately: we don't really need to know how they interact.

That may sound like a cop-out, but it really isn't.

Consider, first, a different question. How exactly do physical objects causally interact among themselves, in purely physical exchanges? The short answer is—no one really knows. We may invoke the laws of motion to explain, but of course no one really understands why those laws take precisely the form they do or how they work.

The point is this. In any investigation of the purely physical world you reach a point where you must admit certain "primitives": facts or phenomena for which no further explanation is possible. You may explain why bodies move the way they do by invoking the laws of motion, but then you simply accept the laws of motion as a basic fact about how the universe is made.

Descartes's view about mind-body interaction was precisely the same: that mental states and events can cause physical states and events, and vice versa, is a primitive feature of the world. If accepting "primitives" is good enough in physics, then it should be good enough for mind-body interactions as well.

Conclusion

Descartes recognized, as we noted, that mental phenomena have a qualitative aspect, or qualia.

But as we have seen, qualia are a two-edged sword: they provide the strongest arguments for Dualism in the first place, but understanding precisely how they fit into the physical world remains perhaps the most impenetrable part of understanding the mind-brain relationship. To this day, no one understands the nature of qualia and their relationship to the brain.

So Descartes was way ahead of his time in recognizing that qualia simultaneously provide both the best evidence, and the biggest difficulty, for Dualism.

FOR GREATER UNDERSTANDING

Questions

1. Both Malebranche and Leibniz offer rather surprising theories of the causal relations between minds and bodies, in which God plays a significant role. If you are a theist, which of their systems strikes you as more plausible, and why? If neither is plausible, then why not?
2. Of the three theories in play in this chapter—those of Descartes, Malebranche, and Leibniz—which one, in your opinion, could best accommodate the existence of free will, and why?

Suggested Reading

Descartes, René. "Meditations on First Philosophy (1641), Meditation 6" and "Passions of the Soul (1649)." *Philosophy of Mind: Classical and Contemporary Readings*. Ed. David J. Chalmers. New York: Oxford University Press, USA, 2002.

Other Books of Interest

Leibniz, G.W. "A New System of the Nature and Communication of Substances, and of the Union of the Soul and Body (1695)." Trans. Roger Ariew and D. Garber. *Modern Philosophy: An Anthology of Primary Sources*. Eds. Roger Ariew and Eric Watkins. 2nd ed. Indianapolis, IN: Hackett Publishing Company, 2009.

Malebranche, Nicolas. *The Search After Truth* (1674–75), VI.II.3. Trans. Roger Ariew and Marjorie Grene. *Modern Philosophy: An Anthology of Primary Sources*. Eds. Roger Ariew and Eric Watkins. 2nd ed. Indianapolis, IN: Hackett Publishing Company, 2009.

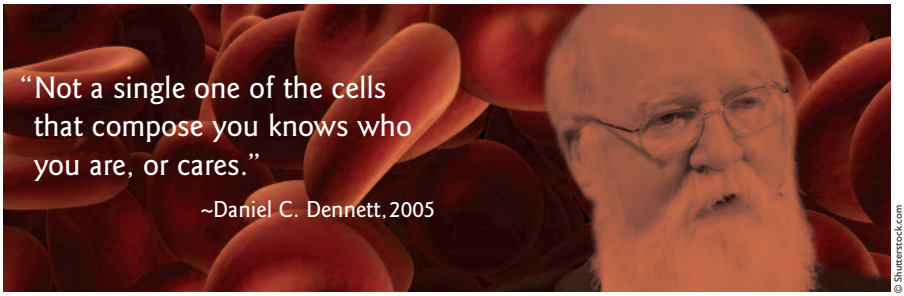
Websites of Interest

1. The *Early Modern Philosophy* website provides the complete text of Nicholas Malebranche's *Dialogues on Metaphysics*. — http://www.earlymoderntexts.com/f_maleb.html
2. The *Stanford Encyclopedia of Philosophy* provides an entry on G.W. Leibniz titled "Leibniz's Philosophy of Mind" (first published September 22, 1997, with a substantive revision October 9, 2007). — http://plato.stanford.edu/entries/leibniz-mind

Lecture 8

What Are Thoughts?

The Suggested Reading for this lecture is Daniel C. Dennett's "True Believers: The Intentional Strategy and Why It Works" in *Philosophy of Mind: Classical and Contemporary Readings*, edited by David J. Chalmers, pp. 556–568.



Introduction

We're now ready to focus on "intentionality," the property of meaningfulness or "aboutness" displayed by many mental states. Some philosophers think that intentionality is precisely what distinguishes the mental from the physical, but we've questioned whether intentionality is an entirely *adequate* criterion for that purpose. There seem to be physical things, for example, which possibly possess intentionality as well: linguistic objects, photographs, thermostats, the electrical circuits in computers. Any theory of intentionality will have to say something about these.

Physicalism and Intentionality

We'd love to find a *Physicalist* theory of intentionality. Such a theory would explain how intentionality can arise in purely physical systems. That would be, in effect, to solve the mind-body problem, or at least a large part of it.

In this lecture, we'll look at the innovative theory of Daniel Dennett, who tries to illuminate intentionality without attempting an explicitly Physicalist account of it. In Lecture 9 we'll look at an influential Physicalist theory of intentionality, the "Computational Theory of Mind."

Original versus Derived Intentionality

Recall those physical things that seemed to be candidates for intentionality: linguistic objects, photographs, and thermostats. Let's now briefly explore whether we may be able to resist seeing these as genuinely possessing intentionality.

The way to do this is to invoke a distinction between "original" intentionality and "derived" intentionality. Many things may be treated as *if* they possessed intentionality without actually possessing intentionality. Such things have "derived" intentionality: they derive their intentionality from the beings who treat them as if they were intentional. In contrast, other things—such as our minds—truly have intentionality, in their own right, intrinsically. Thus we have "original" intentionality.

So consider linguistic objects such as words or sentences. Sure, the word "banana" is about something: bananas. But it doesn't have this aboutness intrinsically, on its own. Rather, human beings have decided to use that word to be about bananas. So linguistic objects only have a derived intentionality. To the contrary, our ability to have *thoughts* about bananas does not depend on or derive from anything else; we have original intentionality.

And then there are computer circuits. Keep in mind John Searle's Chinese Room thought experiment from Lecture 6. Computers do all sorts of things, but none of it is meaningful intrinsically. That was Searle's point in insisting that purely formal symbols don't have meanings. What misleads many people about computers is that we project our own intentionality onto them: we interpret the symbols; we provide their meanings. A properly programmed computer possesses at best derived intentionality, but we with minds are possessed of genuine, original intentionality.

Our main question thus becomes: what is the nature of *original* intentionality?

What Makes It the Case That We Have Intentional States?

Consider some simple belief you may have; say, that today is Thursday. Now ask yourself, what makes it the case that you have this very belief? You might answer that some feature or aspect of your brain makes it true that you have this belief. If this is the right sort of answer, then perhaps for most beliefs there's an absolute answer to whether you do or do not have that belief: you have it if the right sort of thing is inside your head, otherwise not.

On this view, the question of whether a person has a particular belief is an objective internal matter of fact. But according to Dennett, that question seeks not an objective answer, but is rather more like the question of whether a person is "immoral," or has "style," or "talent," or would make a "good" spouse: something subjective, a matter of some interpretation.

Dennett refers to his theory as *Interpretationism*: whether an individual has a particular belief is a matter of interpretation. He writes:

> What it is to be a true believer is to be an intentional system, a system whose behavior is reliably and voluminously predictable via the intentional strategy (Chalmers, 2002, p. 557).

Intentional States and Behavior

Under what conditions does a being count as having genuine intentional states? If its behavior is well predictable by *interpreting* it as having intentional states. According to Interpretationism, the attribution of intentional states is an instrument we use to predict behavior, and the "reality" of intentional states is entirely exhausted by this predictive role. Nothing more needs to be said about what's going on "inside" the being.

This may sound strange, but perhaps it isn't. After all, for millennia human beings have been successfully ascribing intentional states to each other while knowing nothing of the inner workings of the brain. But then our basic attribution of intentional states must not be making any reference to those workings.

So don't think of intentional states as "inner states" of a being—they are just tools we use to predict behavior.

But now, what exactly are "tools to predict behavior"?

The Astrological Stance

There are many strategies one might use for predicting someone's behavior. One might be this: take the date and hour and location of the person's birth, and then consult an astrologer to get some predictions about this person's future and future actions. To adopt this strategy is to "take the astrological stance" towards predicting behavior.

The Physical Stance

But there is a better strategy. One takes "the physical stance" towards the prediction of behavior when one treats the being in question as a purely physical being subject entirely to the laws of nature. In principle we imagine we could take this stance towards human beings, though it's obviously limited in practice.

The Design Stance

It is often more convenient to take "the design stance": here one treats the system as if it has some sort of design and will behave pretty much as designed. Consider how we interact with our artifacts: we have a pretty good idea of what they are "designed" to do, and thus can pretty reliably predict what they will do.

But the design stance has its own limits: for many things we may not know the design, or there will be times when things do *not* work as designed. And so, too, the design stance may not always be practical.

The Intentional Stance

Fortunately there's one more tool in Dennett's toolbox: the intentional stance.

> [F]irst you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then you figure out what desires it ought to have, on the same considerations, and finally you predict that this rational agent will act to further its goals in the light of its beliefs. A little practical reasoning from the chosen set of beliefs and desires will in many—but not all—instances yield a decision about what the agent ought to do; that is what you predict the agent *will do* (Chalmers, 2002, p. 558).

Suppose Fred tells you that he'll meet you at the theater at 7:30 PM tonight. Under normal circumstances, you can predict that Fred will be at the theater around 7:30 PM. But how exactly do you make this prediction?

Step One

You assume that Fred is a rational agent.

Step Two

You theorize about which intentional states he has. Beliefs about geography: beliefs about where he is, where the theater is, how to get from here to there. His beliefs about you, that you desire to see this performance, that you desire to meet him on time. Other relevant beliefs, such as that friendship is valuable, and that punctuality is important to friendship, and so on. And there are numerous relevant desires as well.

Step Three

You then do some "practical reasoning": given his beliefs and desires, he'll rationally conclude that he should try to get to the theater at 7:30 PM.

Step Four

You then predict that Fred *will* do what he judges to be most rational to do.

This simple activity turns out to involve a lot of theoretical reasoning. And it is in doing precisely this that we take the intentional stance towards others.

Note, of course, that you do all this while having no knowledge of the inner physical workings of Fred.

The Limits of the Intentional Stance

But of course the intentional stance isn't flawless. Sometimes people are not perfectly rational; sometimes your theory about Fred's intentional states might be wrong. And even when it works it doesn't work precisely. The crucial point is just that the intentional stance works in general, adequately, most of the time.

The Bold Move

The other crucial point is the controversial one: Dennett isn't merely claiming that we use the intentional stance to predict each other's behavior. He's claiming that being "reliably and voluminously" predictable by it is *all there is* to truly having intentional states.

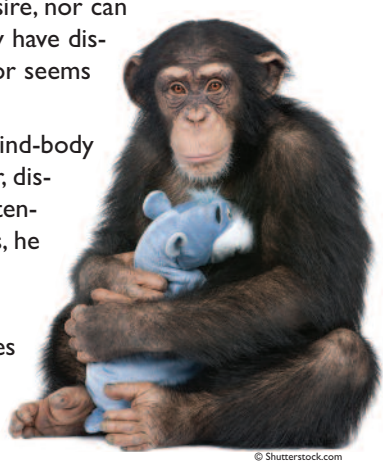
Problems for Dennett's "Intentional Stance" Theory

1. The theory is non-Physicalist: it doesn't attempt to reduce intentional states to something non-mental. So for many, Dennett hasn't explained what most needs explaining.
2. Many thinkers resist Dennett's starting point: that having intentional states is a matter of interpretation.

   For one example, you and Freud would very likely disagree on what beliefs to attribute to your mother, even if you make similarly accurate predictions of her overall behavior. The natural question is, who's right? But Dennett's theory offers no way to resolve such disputes, and that's what seems problematic.
3. On Dennett's theory, whether a being counts as having genuine intentional states may be a matter of degree. You may be convinced that we human beings genuinely have mental states. But what about chimpanzees? And dogs, and snails, and insects, and amoebae? To varying degrees the behaviors of all these animals seem reliably predictable by the intentional stance, but at what point does a creature cross the line into having genuine intentionality?

   Or to put it differently: predictability is a matter of degree, but intentionality does not seem to be. Dennett's theory blurs the line between derived and original intentionality, and that's where many feel it runs into trouble.
4. Dennett simply offers no account at all of *why* the intentional stance actually works. He thus does not and cannot say what makes some particular state of yourself count as a belief or a desire, nor can he explain just why we act "as if" we truly have distinct intentional states. So something major seems to be left out here.

   Moreover, this means he has to give up mind-body causation. For causes always are particular, discrete things. Since Dennett denies that intentional states correspond to internal states, he cannot say that you acted *because* you believed and desired a particular thing.
5. Finally, we know our own intentional states on the basis of introspection—not on the basis of the intentional stance. We know



© Shutterstock.com

in our own cases that whether or not we have such states is not a matter of interpretation, or of degree: it is a perfectly objective truth. But then having intentional states is not constituted by one's predictability by means of the intentional stance.

Conclusion

We began our examination of intentionality by looking at Daniel Dennett's Interpretationism, the theory that having intentional states is a matter of being predictable by the intentional stance. But we also raised some concerns about it, suggesting that it does not quite capture everything we seek to understand about intentionality. In particular, it fails to explain just how intentionality can arise in a purely physical system. We thus turn our attention to the next theory, which lays the groundwork for just that, in the next lecture.

FOR GREATER UNDERSTANDING

Questions

1. How convincing is the distinction between "original" and "derived" intentionality? Or to put it differently, would we be justified in denying *real* intentionality to things like photographs, thermostats, and computers, merely because (say) they lack consciousness?
2. Is intentionality a matter of degree, as described in the third problem for Dennett's theory? If it isn't, could Dennett's theory accommodate some "cut-off" point where a being transitions directly from lacking intentionality to having it?

Suggested Reading

Dennett, Daniel C. "True Believers: The Intentional Strategy and Why It Works." *Philosophy of Mind: Classical and Contemporary Readings*. Ed. David J. Chalmers. Pp. 556–568. New York: Oxford University Press, USA, 2002.

Other Books of Interest

Dahlbom, Bo. *Dennett and His Critics: Demystifying Mind*. Malden, MA: Blackwell Publishers, 1995.

Goldberg, Sanford, and Andrew Pessin. Chapter 2: "The Mind-Body Problem." *Gray Matters: An Introduction to the Philosophy of Mind*. Armonk, NY: M.E. Sharpe, 1997.

Ross, Don, Andrew Brook, and David Thompson, eds. *Dennett's Philosophy: A Comprehensive Assessment*. Cambridge, MA: MIT Press, 2000.

Websites of Interest

1. Daniel C. Dennett's home page at Tufts University provides a short biography, recent published articles, and links to interviews and other information about the philosopher. — http://ase.tufts.edu/cogstud/incbios/dennettd/dennettd.htm
2. *The Guardian* (London) website provides an article by Andrew Brown from 2004 about Daniel C. Dennett entitled "The Semantic Engineer." — http://www.guardian.co.uk/books/2004/apr/17/peopleinscience.philosophy

Lecture 9

The Language of Thought

The Suggested Reading for this lecture is Jerry A. Fodor's "Propositional Attitudes" in *Philosophy of Mind: Classical and Contemporary Readings*, edited by David J. Chalmers, pp. 542–555.

Introduction

There are two components of intentionality that need examination: the thing that is "about" something, and the something this thing is about. In this lecture we shall focus on the former, building on Functionalism from Lecture 5 as we explore "The Computational Theory of Mind."

The Computational Theory of Mind: Syntax and Semantics

The basic idea of the Computational Theory of Mind (CTM) is that the mind is like a computer. Computational states have two main aspects, their "syntax" and "semantics." "Syntax" refers to their physical properties, to the state or entity that itself is occupying the relevant causal or functional role. "Semantics," ultimately, refers to the role itself, and thus refers to the content of the state, its "meaning."

The idea is that, in a computer, a causal network of purely physical states mirrors or instantiates a network of meanings.

To illustrate, imagine two nodes each connected to a third node. They are set up so that if, and only if, both of the first two nodes are lit up, then the third node also lights up. This simple physical system represents the meaning of the word "and." Allow the first node to represent a particular proposition; the second to represent a different one; then the third node will represent the assertion that both the first one *and* the second one are true.

In this example, the purely physical causal aspects of the system are its syntax. What each of these elements represents or stands for is the semantics of the system. The idea is that you can establish physical systems inside computers

with a certain syntax that can represent particular meanings—much as these nodes can represent the respective propositions. And the idea of CTM is that the brain is similar to a computer in just those respects: its physical properties constitute a syntax, but its overall activity also instantiates a semantics.

The most influential version of CTM is from Jerry A. Fodor, and it is known as "The Language of Thought" (LOT).

The Language of Thought

LOT is the theory that thought itself, or intentionality in general, has a linguistic structure.

To say this is to say that the nature of intentional states is very much like sentences. How so? Sentences are complex entities composed of simpler entities, namely, words. But there are very specific rules that govern which combinations of words count as legitimate sentences. "The creature ate the green eggs and ham" is a legitimate sentence in a way that "The ate creature the eggs ham and green" is not. These rules are often called the syntax of the language. And of course in addition to the syntax of a language there is the semantics: what the words mean, and what the sentences composed of them mean. In general, the meaning of the sentence is a function of the meanings of the words in conjunction with the syntax of the sentence.

So ordinary language is characterized by its syntax and its semantics. And according to LOT, so too are intentional states, themselves complex entities composed of simpler entities ("concepts") combined according to various syntactic rules. And they too have a semantics, determined by the meanings of their parts along with the rules of composition.

Two Important Points About LOT

LOT's primary contribution to the Physicalist project is with respect to clarifying the mental rather than connecting it to the neural. Ultimately neuroscience must seek what brain activities could subserve the linguistic structures of thought.

Secondly, while LOT claims that thought has a linguistic structure, the language of thought—the language in which we constitute our intentional states—cannot be identified with any particular ordinary language.

To see why, consider the fact that you and a French speaker can think the same thought: you can both think that it's raining. But you can do this even if you don't share an ordinary language. But then the thought itself may not be identified with either ordinary language. So the language of thought is neither French nor English, nor any ordinary language.

Fodor calls it "mentalese." Mentalese is what perhaps all humans share, insofar as we are capable of intentionality—even if there is no single ordinary language we share.

Which Leads to a Fascinating Argument for LOT

Basically, Fodor argues, you need to already have a language in order to learn a new language. Consider how we typically learn a second ordinary language. We hear certain words and we formulate hypotheses about what those words mean. Learning a new language involves using our original language to *formulate* these hypotheses. So we use the language we have in order to learn a new language.

But now consider how a child learns her *first* language. The process would seem to be the same. She hears certain words and formulates hypotheses about what they mean. But she doesn't already have an ordinary language in which to formulate those hypotheses. So she must have another language, a more basic language—a language of thought.

More Reasons to Believe in the Language of Thought

Fodor argues that accepting the existence of mentalese explains many important features or properties of thought.

Systematicity

To say that language is systematic is to say this: if a given combination of words constitutes a legitimate and meaningful sentence, then so too will other combinations of words that are related to that combination in systematic ways. So, for example, if "Fred loves Jane" is legitimate and meaningful, then so will be "Jane loves Fred."

The reason language is systematic is straightforward: sentences are composed of words according to syntactical rules, and if the rules allow certain word combinations, they'll allow other similar combinations.

But the same now holds for thought itself. If you're capable of *thinking* that X loves Y, you're also capable of *thinking* that Y loves X. And if thought is structured linguistically—consisting of parts combined according to rules—then we can offer the same explanation of thought's systematicity as we did of language's.

So systematicity is evidence for a language of thought.

Productivity

We are capable both of creating and of understanding absolutely brand new sentences that we have never encountered before, and doing so nearly instantly. How can that be? Because language is "productive." To say that is to say that new sentences can be built from old parts by combining them, via the rules, in new ways. Since the new sentences are built from familiar parts via familiar rules, it's a fairly simple matter for us to create them and to understand them.

The same is true for thought itself. We are capable of thinking new thoughts apparently without limit, and to think new thoughts about the same things.

And if thought has a similar structure to language, we can understand precisely why. The productivity of thought is therefore evidence that thoughts are composed of recombinable parts, and thus that we have a language of thought.

Empirical Considerations

Finally, as Fodor puts it, belief in LOT simply fits best with what cognitive psychology actually does. For just one example, our ability to recognize, and to respond appropriately, to linguistic ambiguity seems nicely explainable if we accept the existence of mentalese. A word will count as ambiguous, in brief, if it is mapped onto two distinct concepts. That suggests, first, that we are capable of thinking thoughts that are distinct from the words we use to express them. It suggests, second, that we can swap distinct components in and out of the same linguistic structure, as we do when we grasp the distinct meanings of an ambiguous sentence. But to say both these things is to say that thought itself has a linguistic structure, that is, we have a language of thought.

Four Criticisms of LOT

1. LOT is limited in several ways. For example, it only addresses intentionality, and says nothing about qualia. Nor does it give us everything we need with respect to intentionality, for we still don't understand what makes the components of thought, the concepts *themselves*, have a meaning or content.
2. Nativism: "Nativism" is the doctrine that certain mental things, such as ideas, or concepts, or even knowledge, count as "innate." To be a nativist about something is to claim that the thing is already built into the mind from birth in some sense. LOT strongly suggests that mentalese would be innate. And while the debate over nativism in general is beyond our scope, some raise this as a strike against LOT.
3. Mental causation: The worry here is this: any causing that an intentional state does will be entirely a matter of its syntax or physical properties. The fact that these structures also have a semantics will be causally irrelevant. That in turn suggests that what is specifically *mental* about these states, their intentional content, is causally irrelevant. And that of course is to deny that the mind is causally relevant.

4. Mind and language? Finally, the most general concern about LOT is that it simply gives too much emphasis to the fact that we typically express our intentional states by using sentences. To many thinkers, intentionality, and cognitive life more generally, often involves much more than language—and indeed there is a long tradition in philosophy that holds that what we can express in words or sentences is but a pale reflection of the way our mind operates.

   Or to put this succinctly: a picture is worth a thousand words because our perceptual experience goes well beyond what can be captured in words. But then we should be skeptical that our perceptual experience, our intentionality in general, has a genuinely linguistic structure.

Conclusion

In this lecture we briefly sketched the Computational Theory of Mind, and then focused on Jerry Fodor's Language of Thought hypothesis. We explored the idea that intentional states have a linguistic structure, leaving it open, for this lecture, exactly how the components of our intentional states, and thus ultimately the intentional states themselves, acquire their contents or meanings. It is to that topic that we next turn.

FOR GREATER UNDERSTANDING

Questions

1. The most interesting argument for Fodor's "Language of Thought" is based on the claim that one requires a language in order to learn a language. While that may be true for the way we learn second languages, is it really true for the way we learn our first ordinary language? Do we really "formulate hypotheses" in a "language of thought" when we first learn to speak? If not, how *do* we learn language?
2. Does our perceptual experience really transcend language, as one of the criticisms of LOT asserts? If so, does that mean that Fodor's theory must be wrong, or could it accommodate that fact?

Suggested Reading

Fodor, Jerry A. "Propositional Attitudes." *Philosophy of Mind: Classical and Contemporary Readings*. Ed. David J. Chalmers. Pp. 542–555. New York: Oxford University Press, USA, 2002.

Other Books of Interest

Fodor, Jerry A. *LOT 2: The Language of Thought Revisited*. Oxford, UK: Oxford University Press, 2008.

Loewer, Barry, and Georges Rey, eds. *Meaning in Mind: Fodor and His Critics*. Malden, MA: Blackwell Publishers, 1992.

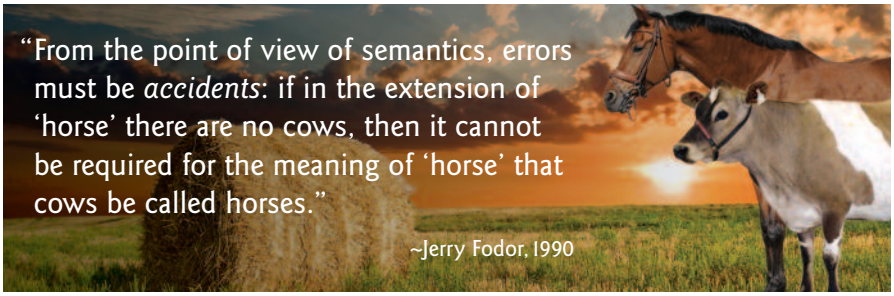
Websites of Interest

1. The *Stanford Encyclopedia of Philosophy* provides an article by Murat Aydede entitled "The Language of Thought Hypothesis," which explores Jerry Fodor's philosophy on the topic in detail. — http://plato.stanford.edu/entries/language-thought
2. A *Notre Dame University* review (with commentary) by Mark Wilson (University of Pittsburgh) of Jerry Fodor's book *LOT 2: The Language of Thought Revisited*. — http://ndpr.nd.edu/review.cfm?id=15366

Lecture 10

Theories of Meaning

The Suggested Reading for this lecture is Jerry A. Fodor's *A Theory of Content and Other Essays*, "A Theory of Content, I: The Problem" (chapter 3, pp. 51–88) and "A Theory of Content, II: The Theory" (chapter 4, pp. 89–136).



Introduction

In this lecture we aim to get to the heart of intentionality by asking this question: How do the elements of thought come to be meaningful, to be about things? By virtue of what does some entity in our minds (or brains) come to be about something?

Theory One: The Resemblance Theory

A very common idea from seventeenth-century philosophy was that what we perceive are not physical objects themselves but rather our own "ideas" of them, our own mental images or representations of those objects.

This surprising theory was supported by many arguments. Perhaps most compelling was the "dream" argument. During a dream, most agree, what you perceive are mental images. But ordinary waking perception is qualitatively indistinguishable from dream perception. So we should conclude that even during ordinary perception what we perceive are mental images.

This leads us back to the question: By virtue of what does the mental image come to represent the thing it represents?

The Resemblance Theory answers that mental objects are about things in the world by virtue of resembling them.

Problems for the Resemblance Theory

1. For Physicalists the theory won't get off the ground. For a Physicalist, mental representation occurs in the brain, but brains don't resemble anything other than other brains. So by the Resemblance Theory no brain could represent anything other than some other brain.

2. But even the non-Physicalist has concerns.

   Two things resemble each other if they share properties. But then it's not clear just how "images" would count as literally resembling the thing they are images of. A photograph of a cat, for example, shares very few properties with cats; and *mental* images would have even less in common with real physical objects. Mental images could not represent physical objects, in short, by resembling them.

3. There are abstract things we can think about that *nothing* could possibly resemble: truth, virtue, democracy. So resemblance can't explain our ability to think about them.

4. Finally, language represents without "resemblance." But now similar considerations apply to the words of mentalese, the language of thought. Indeed, thought has linguistic features that images just cannot capture. In particular, intentional states are discrete while images are rich in information. Suppose you think "Fred has short hair." That is a thought about only Fred and his short hair. But any image representing this would represent many of Fred's aspects besides his hair. No image could represent just *that one single thought*.

Theory Two: The Causal Covariance Theory

Our next theory replaces mental images with the "symbols" of the Computational Theory of Mind. These purely syntactic entities are said to be meaningful, but not by virtue of resembling what they're about. So how?

The Basic Idea

A particular mental representation is about something insofar as it "causally covaries" with the thing. To illustrate, imagine a certain squiggle in one's brain—a pattern of neural activity. What makes that squiggle, that symbol, be about cows, for instance, we might ask? The answer is that there is some regular causal connection between the presence of cows and the activation of that symbol. No resemblance between squiggle and cows is necessary, as long as the causation is in place.

The First Version

"The squiggle means *cow* if and only if all and only cows cause the squiggle to be activated."

But several problems arise.

1. *The "breadth problem."* We want the squiggle to mean *cow*, to refer to all cows in general—but we surely can't require that all cows actually succeed in causing the squiggle to be activated in the agent's head.

2. *The "depth problem."* Suppose that the squiggle is caused every time any cow is present in the agent's perceptual field. Would that make the squiggle mean *cow*? Well, the causal chain between the actual cow and the activation

of the symbol actually has many members. The cow causes light to reflect; the light causes eyeball activity; the eyeball activity causes brain activity, which then results in the activation of the squiggle. If the squiggle is regularly caused by actual cows, then it will also be regularly caused by all those intermediaries. But then why does the squiggle mean *cow* and not any or all of those other things as well?

3. *The "non-existent objects problem."* We can think about non-existent things. Yet how can a squiggle in our head be caused by something that doesn't exist?
4. *The "misperception problem."* We sometimes *mis*perceive things. A quick glance at a horse at dusk may cause us to think *cow*—but then those are cases where something *other* than a cow causes that squiggle to get activated. Then it's not the case that "only" cows cause the symbol.

The Second Version

Unfortunately things get very technical as philosophers try to revise the Causal Covariance Theory to meet these problems, so we'll have to satisfy ourselves with a quick sketch of the attempted remedy. The key revision is that we should express the causal relation in terms of "causal laws." Rather than relating squiggles and cows, we might say, "The squiggle means *cow* if and only if there is a causal law of the form 'cows cause squiggles.'"

One crucial consequence of invoking laws here is that it allows us to say, for example, "If there *were* a cow present, then it *would* cause the squiggle to be activated."

This version of the Causal Covariance Theory might handle some of the earlier problems. We can now say that the squiggle applies to all cows because all *would* activate the squiggle, *were* they to be around; that handles the "breadth problem." Similarly an object *would* cause the squiggle *were* it to exist; that handles the "non-existing objects problem." And, arguably, the revised theory can handle the "depth problem," though we won't go through the details here.

But what about the "misperception problem"?

The Problem of Misperception and the Language of Thought

The problem, again, is this. Sometimes a horse at dusk might cause us to think "cow." But then our squiggle, which covaries with cows, might covary even *more* reliably with a "cow-or-horse-at-dusk." But then the theory would dictate that the squiggle means not *cow*, but rather *cow-or-horse-at-dusk*. In which case we haven't actually misperceived at all—for what we saw, here, was indeed a cow-or-horse-at-dusk.

What we want to say, however, is that a horse-at-dusk might sometimes *mistakenly* cause us to think *cow*. We want our squiggle to mean just *cow* even though things other than cows can cause it. And that's where the Causal Covariance Theory goes wrong.

Fodor's Asymmetric Dependence Theory

In solving the problem of misperception, Jerry Fodor observes that what's needed is some way to distinguish cases where cows *correctly* cause squiggles and horses *mistakenly* cause them. If we can break the apparent "symmetry" between such cases, then we could restrict the squiggle's meaning to the correct cases while excluding the incorrect cases.

What we must note is the actually *a*symmetric way in which the mistakenly activated squiggles depend on correctly activated ones. Very simply: a squiggle wouldn't activate on seeing a horse-at-dusk unless it were *already* the case that the squiggle means *cow*. But the reverse is not true: having that squiggle activate on seeing a cow does *not* depend on the fact that sometimes horses-at-dusk can cause that squiggle.

Thus, Fodor concludes, horse-caused squiggles are dependent on cow-caused squiggles but not vice versa. So cow-caused squiggles have a kind of priority over the horse-caused squiggles in determining the symbol's meaning. We'll conclude our discussion of Fodor here.

A Shocking New Way of Thinking About Intentionality

The problem of intentionality is clearly a difficult one. And surely the bombshell that Harvard philosopher Hilary Putnam dropped into the discussion in 1975 didn't help matters much. It took the form of a fascinating thought experiment that suggests the quite shocking conclusion that, put bluntly, our intentional states are not actually "in our head"!

But first some background.

Internalism versus Externalism in the Philosophy of Mind

Intuitively, our thoughts are insulated from the external world. What thought we're thinking at a given moment seems to be entirely a function of what's "inside" us, even if our thoughts may often be caused by external things.

One way to probe this intuition is to imagine a molecule-by-molecule duplicate of your brain. Such a brain, it seems intuitively, would have exactly the same mental states as yours. To agree with this intuition is to think of mental-states as being "in the head": just which mental state one is in depends only on what's inside the head.

The doctrine that this is so is known as "Internalism."

But Now Some Science Fiction

Daniel Dennett begins chipping away at Internalism by asking whether a molecular duplicate of Dennett himself would truly have the same mental states as original-Dan. No, he answers—for original-Dan has thoughts about his wife Susan, but duplicate-Dan has never met Susan. And similarly for many other mental states.

If that is so, then brain duplicates need *not* be mental-state duplicates. The doctrine that *this* is so is called "Externalism."

More Science Fiction: Twin Earth

Hilary Putnam's thought experiment imagines not just a molecular duplicate of some individual person, but of Earth as a whole.

There is only one difference between Earth and Twin Earth. On Earth, the stuff we call "water" is made of H_2O; but on Twin Earth, the stuff they call "water," which is indistinguishable from H_2O, is something else—let's say "XYZ."

Suppose you think to yourself, "I sure would like some water." At the same time Twin You thinks, "I sure would like some water." Are You and Twin You thinking the same thought?

Perhaps not, despite your identical brain states. For intentional states are defined by their "aboutness," by *what* they are about. Thoughts about different things count as different thoughts—and yours is about H_2O, while your twin's is about XYZ. Therefore, You and Twin You are *not* having the same thought.

To use Putnam's famous expression—intentional states are not simply "in the head" after all.

The Aftershocks

If Putnam's "Externalism" is correct, if what mental state you are in is partly a function of what's *external* to your brain, then the following apply.

1. We can no longer treat mental properties as simply depending on what's going on in the brain.
2. We might have to reject the old idea that we have a kind of privileged access to our own mental states.
3. The problem of mental causation gets even thornier, for if the intentional content of your mental state is determined by factors outside your brain altogether, then it's hard to see how intentional content can be causally relevant to your behavior.

Conclusion

In this lecture our discussion of intentionality reached its peak. We explored two different theories about how entities in the brain or mind might come to be "about things"—the Resemblance Theory and the Causal Covariance Theory. We discovered that what our intentional states are about might, in the end, depend on factors outside our brains—which surely complicates our attempt to understand the relationship between mind and brain. And overall we can now appreciate just how difficult the problem of intentionality is—and that there is no clear solution to it in sight.

FOR GREATER UNDERSTANDING

Questions

1. It's claimed in this lecture that photographs do not, strictly speaking, resemble the objects of which they are photographs. Is that really true, or does it merely mean that we need some other definition of "resembling" than the one given in the lecture? If so, what definition would work?
2. Do you agree with the Twin Earth experiment's conclusion that two individuals with exactly the same brains and brain states might nevertheless count as being in different intentional states? If so, then does that support some version of Dualism? But if not, then how could we describe the intentional state the two Twins allegedly share, since they seem to have different objects?

Suggested Reading

Fodor, Jerry. *A Theory of Content and Other Essays*. "A Theory of Content, I: The Problem" (chapter 3, pp. 51–88) and "A Theory of Content, II: The Theory" (chapter 4, pp. 89–136). Cambridge, MA: MIT Press, 1992.

Other Books of Interest

Goldberg, Sanford, and Andrew Pessin. Chapter 3: "Theories of Mental Content." *Gray Matters: An Introduction to the Philosophy of Mind*. Armonk, NY: M.E. Sharpe, 1997.

Putnam, Hilary. "The Meaning of 'Meaning.'" 1975. Reprinted in *The Twin Earth Chronicles: Twenty Years of Reflection on Hilary Putnam's "The Meaning of 'Meaning'."* Eds. Andrew Pessin and Sanford Goldberg. Armonk, NY: M.E. Sharpe, 1996.

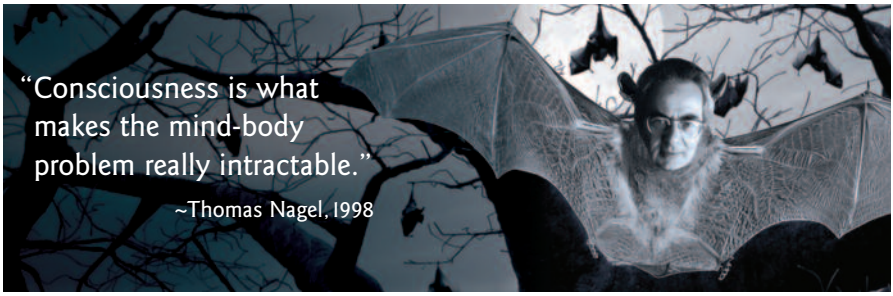
Websites of Interest

1. *ReVEL* is a journal based in Portugal that dedicates each issue to a specific field of linguistic studies. Volume 5, no. 8, 2007, featured an interview with Jerry Fodor entitled "Semantics: An Interview with Jerry Fodor" (select option for English-language site). — http://www.revel.inf.br
2. *BloggingHeads TV* website features a video dialogue between Jerry Fodor and fellow philosopher Elliott Sober (University of Wisconsin-Madison) entitled "Who Got What Wrong." — http://bloggingheads.tv/diavlogs/26848

Lecture 11

Consciousness and the Physical World, Part I

The Suggested Reading for this lecture is Thomas Nagel's "What Is It Like to Be a Bat?" in *Philosophy of Mind: Classical and Contemporary Readings*, edited by David J. Chalmers, pp. 219–226.



Introduction

In this lecture we turn to qualia, or "consciousness." What is the nature of the phenomenon, how can we explain it, and how can it be reconciled with the physical nature of our brains?

Unlike intentionality, consciousness is quite difficult to define. Nor do many feel that the problem of consciousness is even remotely close to being solved. Indeed, the problem seems intractable to many. If there is any good evidence for Dualism, it is the existence of consciousness.

Let's begin by clarifying what philosophers mean when they talk about consciousness.

"Phenomenal" Consciousness

The modern use of the phrase "phenomenal" traces back to Immanuel Kant (1724–1804), who distinguished between "phenomena" and "noumena." The word "noumena" referred to objects and events as they are "in themselves," independent of the mind's attempt to know them. In contrast, "phenomena" referred to the very same objects and events insofar as they *are* knowable or known. What drives Kant's distinction here is recognition that the nature of our minds might influence how we perceive and understand these objects—and so how things are perceived by us, or how they "appear" to us, might not be the same as how they are "in themselves."

When they speak of consciousness philosophers often mean "phenomenal consciousness": the nature of our awareness of things, which will be partly a function of the structure of our own minds.

Appearance and Reality, and Qualia

So characterized, the study of consciousness is related to the classic distinction between appearance and reality.

Again, how things "appear" to us can often be different from how things "really are." When philosophers address consciousness, they are interested primarily in the former. Even ordinary cases of reliable perception involve "appearances," in all the sensory modalities. How blue appears or looks is different from how red looks; how garlic appears or smells is different from a pineapple; and so on. Everything we sense "appears" to us, has a "way" that it appears, or is experienced by us in a certain way; there's "something that it's like" to experience the thing.

Moreover, "how things appear" may depend crucially on the individual, or the species. How colors look to humans may differ from how they look to other creatures; indeed, how a given object's color looks to one person may well vary from how it looks to another. How things appear in these cases is very much a "subjective" matter.

Our conscious experience of the world, then, has a qualitative character, a way that things feel or seem or appear, which is highly subjective. Philosophers refer to this as the "qualia" of mental states. And the "problem" is whether the existence of consciousness, of qualia, can be accommodated within a purely Physicalist conception of the world.

Thomas Nagel on "What Is It Like to Be a Bat?"

In 1974, Nagel published the article "What Is It Like to Be a Bat?" Strictly speaking, Nagel is arguing not that Physicalism is false, but that we have no conception of how it could possibly be true. Still, many read his article as an argument against Physicalism, and that is how we will read it.

Nagel proceeds by discussing the "subjective character of experience," or "what it's like to be a given organism." He argues that this is always closely connected with a single "point of view": wherever there is conscious experience there is always some individual subject whose experience it is, who stands in a special relationship *to* that experience.

Purely physical states and processes, to the contrary, lack this kind of subjectivity or point of view. Physicalist theories strive for "objectivity," for truths that are independent of any point of view or perceptual system. Brain processes, for example, are paradigm examples of objective phenomena.

But that's the problem: conscious experience is subjective, while purely physical processes are objective. But how can one get subjectivity from a series of objective things? Hence we have no way of explaining how consciousness derives from physical phenomena.

The Brain in the Bat

Nagel illustrates the preceding argument with his famous bat example:

> . . . [T]he essence of the belief that bats have [conscious] experience is that there is something that it is like to be a bat. Now we know that most bats . . . perceive the external world primarily by sonar, or echolocation, detecting the reflections, from objects within range, of their own rapid, subtly modulated, high-frequency shrieks . . . [T]he information thus acquired enables bats to make precise discriminations of distance, size, shape, motion, and texture comparable to those we make by vision. But bat sonar . . . is not similar in its operation to any sense that we possess, and there is no reason to suppose that it is subjectively like anything we can experience or imagine (Chalmers, 2002, pp. 220ff).

To say a bat has conscious experience is to say that there is something that it is like to be the bat; and in particular, to perceive the world by sonar. But given the very different nature of bats' sonar from our own perceptual systems, can we ever know what it *is* like for a bat to perceive by sonar?

Nothing stops us from learning much that is physical in nature about bats' sonar. Scientists understand in detail how sound waves work, how bat ears work, how the bat's brain processes the sound signals, and so on.

There's just one thing missing: the subjective nature of the process, namely *what it is like* to navigate by sonar.

In short, we've got good reason to believe that there is something it is like to be these animals. And yet the objective physical facts about how animals' brains work won't allow us to grasp those subjective facts or understand how they arise.

And that sounds like the denial of Physicalism.

Some Conclusions

Nagel's argument suggests several possible closely related conclusions.

1. In addition to all the objective physical facts there are about the world, about how physical brains work and so on, there also exist subjective non-physical facts.
2. More strongly, not merely are the subjective facts different in kind from the objective facts, but they cannot even be explained by the objective facts: subjectivity doesn't seem derivable from objective facts.
3. Finally, there are important limits to what we can know about the world and each other. In particular, the subjective facts about other creatures' conscious experiences are unknowable to us. Even more generally, perhaps no individual human can properly grasp what it's like to be any other individual human being either.

Since Nagel's argument anticipates a later much-debated argument presented by Frank Jackson, we'll save various replies to it until after our discussion of Jackson in Lecture 12.

Colin McGinn's "Can We Solve the Mind-Body Problem?"

Like Nagel, McGinn too does not, strictly speaking, deny Physicalism: in fact, he argues *for* it. Nevertheless, he also argues that we cannot discover how the brain is responsible for conscious experience, and so he too provides the tools for an argument against Physicalism.

His argument has two stages:

First, he argues that there exists some physical property of the brain that accounts for consciousness.

He then argues that the manner in which we form concepts leaves us unable to conceive of or grasp that physical property.

Stage One: McGinn's Brief Argument for Physicalism

McGinn:

> Resolutely shunning the supernatural, I think it is undeniable that it must be in virtue of *some* natural property of the brain that organisms are conscious. There just *has* to be some explanation for how brains subserve minds . . . It is implausible to take these correlations [between mental states and brain states] as ultimate and inexplicable facts, as simply brute (Chalmers, 2002, p. 396).

That's pretty much it: some physical property of the brain must ultimately cause consciousness because otherwise it would be a miracle. We'll see more vigorous arguments for Physicalism in Lecture 14.

Stage Two: Why Physicalism Is Inconceivable to Us

Let's consider, McGinn suggests, how we might try to identify the brain property that produces conscious experience.

First, we could attempt to investigate consciousness itself by "introspection." But that seems quite useless in helping us understand how the *brain* produces consciousness, for it amounts to observation only of the consciousness, and not of the link between the brain and consciousness—which simply isn't available to introspection.

So we'll have to explore the brain itself; and here there are two possible ways to proceed.

First, there's "direct observation" of the brain. But when we do that, all we see is wet bloody mushy stuff, nothing remotely resembling consciousness or explaining conscious experience.

Nor are things any different if you imagine observing the brain on a microscopic level. For all we see there are cells such as neurons, filled with fluids

and molecules and lots of electrical activity—but nothing suggestive of consciousness. Nor is anything different when we consider what neuroscientists "see" by the numerous techniques of modern science and medicine. Patterns of electrical activity, of magnetic activity, of blood flow, and so on—but nothing suggestive of consciousness.

Why not? Because the sorts of properties that we can observe this way just aren't the right sorts to explain consciousness. Many are spatial in nature, but, McGinn notes,

> . . . [W]e cannot link consciousness to the brain in virtue of spatial properties of the brain. There the brain is, an object of perception, laid out in space, containing spatially distributed processes; but consciousness defies explanation in such terms. Consciousness does not seem made up out of smaller spatial processes; yet perception of the brain seems limited to revealing such processes (Chalmers, 2002, p. 399).

If we could somehow bracket away our beliefs that brain activity is responsible for mental activity and look at the brain with fresh eyes, so to speak—we'd realize that none of the observable properties of the brain seem to have anything to do with consciousness whatsoever.

That leaves the final possibility: that we explain how consciousness arises from the brain by using some kind of inference or reasoning.

But is that plausible? McGinn notes that

> . . . [C]onsciousness itself could not be introduced simply on the basis of what we observe about the brain and its physical effects . . . Inference to the best explanation of purely physical data will never take us outside the realm of the physical, forcing us to introduce concepts of consciousness. Everything physical has a purely physical explanation (Chalmers, 2002, p. 399).

In other words, all the reasoning we might do about what we observe and measure in the brain, both its causes and its effects, would remain within the realm of the physical. But then no amount of reasoning about the brain itself will ever introduce consciousness.

McGinn's Conclusion

To discover the physical processes responsible for consciousness we must use either introspection, direct observation of the brain, or some kind of indirect reasoning about it—and none of those will yield what we seek. Since McGinn himself accepts Physicalism, he concludes that we simply cannot explain how Physicalism is true.

But for those with no strong inclinations about Physicalism, McGinn's argument gives us good reason to *reject* Physicalism.

Conclusion

Both Nagel and McGinn argue that we cannot understand how the physical brain gives rise to consciousness or qualia.

FOR GREATER UNDERSTANDING

Questions

1. Once scientists have figured out all the physical processes involved in the way bats navigate and feed by means of sonar, is there really anything left over to explain? Is there really "something that it's like" that goes beyond those sorts of physical facts, something that requires some real commitment to Dualism? Or does the fact that this "something that it's like" resists formulation in words suggest that this "something" is nothing that needs any special scientific attention?
2. Both Nagel and McGinn resist the conclusion that Physicalism is false. But don't their arguments very strongly support that conclusion? How exactly can one avoid the conclusion that Physicalism is false if one accepts either or both of their arguments?

Suggested Reading

Nagel, Thomas. "What Is It Like to Be a Bat?" *Philosophy of Mind: Classical and Contemporary Readings*. Ed. David J. Chalmers. Pp. 219–226. New York: Oxford University Press, USA, 2002.

Other Books of Interest

Goldberg, Sanford, and Andrew Pessin. Chapter 7: "Consciousness." *Gray Matters: An Introduction to the Philosophy of Mind*. Armonk, NY: M.E. Sharpe, 1997.

McGinn, Colin. "Can We Solve the Mind-Body Problem?" *Philosophy of Mind: Classical and Contemporary Readings*. Ed. David J. Chalmers. Pp. 394–405. New York: Oxford University Press, USA, 2002.

Recorded Books

McGinn, Colin. *Discovering the Philosopher in You: The Big Questions in Philosophy*. Modern Scholar Series. Prince Frederick, MD: Recorded Books, 2004.

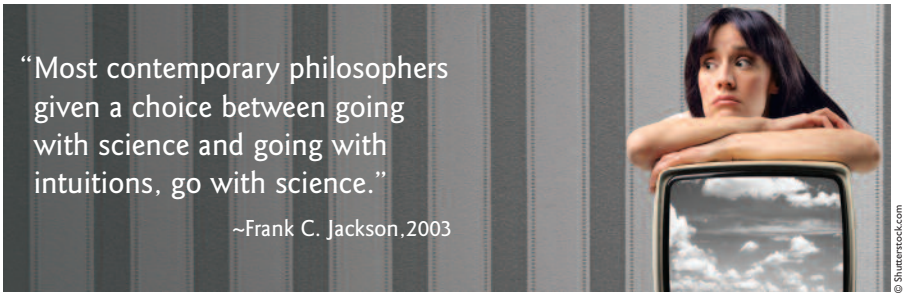
Websites of Interest

1. Professor Thomas Nagel's home page on the New York University website provides links to five of his papers available online. — http://philosophy.fas.nyu.edu/object/thomasnagel
2. Colin McGinn discusses philosophy, beliefs, and other ideas on his blog. — http://mcginn.philospot.com

Lecture 12

Consciousness and the Physical World, Part II

The Suggested Reading for this lecture is Frank C. Jackson's "Epiphenomenal Qualia" and "What Mary Didn't Know" in *Philosophy of Mind: Classical and Contemporary Readings*, edited by David J. Chalmers, pp. 273–280.



Introduction

In this lecture we examine two more arguments for the conclusion that consciousness, or qualia, simply doesn't fit into the Physicalist world.

Metaphysical Background to Saul Kripke's Argument

Saul A. Kripke (1940–) is an American philosopher who was a child prodigy and earned an undergraduate degree at Harvard. He never pursued a graduate degree, but has taught at several institutions, including Princeton (where he is professor emeritus) and City University of New York. Kripke has also lectured at Oxford and has received several honorary doctorates.

Here is a thesis we've seen before: "mental states are identical to brain states." In Lecture 4 we discussed Leibniz's Law, an important criterion for determining identity, which stated that if something X is identical to something Y, then whatever is true of X is also true of Y; or, in reverse, if there's something true of X that isn't true of Y, then $X \neq Y$. What Kripke now offers is some additional insight into the nature of identity.

Kripke's basic claim is this: "Sentences asserting an identity between two kinds of things, if true, express necessary truths."

Necessary versus Contingent Truths

Philosophers distinguish between a sentence's being "necessarily" true, or merely "contingently" true.

To say that it is "necessarily" true is to say that it couldn't possibly be false, that there are no conceivable circumstances under which it would or could be false,

that denying the sentence would lead to some sort of contradiction. Prime candidates for necessary truths are usually mathematical and logical truths.

In contrast, most other kinds of true sentences seem to express merely "contingent" truths. To say that a sentence is contingently true is to say that, though it does happen to be true, it didn't or doesn't *have* to be true, that there are conceivable circumstances under which it would or could be false, that denying it needn't involve any contradictions.

So Kripke's point is this: where X and Y are in fact identical, then they are necessarily identical.

But why should we believe this to be so? It seems pretty easy to imagine cases of merely contingent identity. The Morning Star is identical to the Evening Star, true; but we can conceive of possible cases where what rises in the morning would be different from what rises in the evening, so that identity doesn't seem *necessary*.

Brief Argument for the Necessity of Identity

Kripke responds to that counterexample by noting that it's rightly attempting to generate a case of contingent identity. But the situation just generated is *not* one, he claims, where the Morning Star is not identical to the Evening Star.

Given that they *are* identical, in fact, it is *not* possible to conceive of a circumstance where they aren't identical. To say they are identical is to say that there is one object (Venus) to which we have given two different names. If we focus on that object itself, then it doesn't even make sense to imagine the actual Morning Star—Venus—being different from the actual Evening Star—also Venus. For that would be to imagine a scenario in which Venus is not Venus. And the scenario sketched above is most certainly not *that*.

So the Morning Star *is* necessarily identical to the Evening Star—for Venus is both, and Venus is necessarily identical to itself.

And so, Kripke concludes, identity is necessary after all.

Kripke's Argument Against the Identity of Conscious States to Brain States

With the necessity of identity established, Kripke's main argument is straightforward. If you *can* generate a possible circumstance in which one thing is present without the other, then the two are not identical. And we can do that with respect to qualia and brain states.

We can easily conceive of circumstances in which, for example, the firing of C-fibers—the brain states scientists associate with pain sensations—might occur without the sensation of pain, and vice versa. Just think about brain structures and activities; and think about pain. Nothing about the idea of brain molecules moving suggests anything about feelings or sensations. And nothing about the idea of feelings or sensations suggests anything about physical activities. It's thus easy to conceive of either one occurring without the other.

But then by the necessity of identity, it follows that pain and brain states are not identical, and so that Physicalism is false.

Kripke's Argument and Earlier Issues

Suppose Kripke is right that conscious mental states are separately conceivable from brain states (and thus not identical to them). If so, then there are conceivable scenarios in which either (1) there are very different correlations between the mental and the physical, and/or in which (2) there might be neural events just like our own occurring without any sort of conscious mental states whatsoever.

But now scenarios of the first sort would include cases of "inverted qualia"; and those of the second would include cases of "absent qualia" or "zombies." So Kripke's principle about the necessity of identity ultimately underlies Lecture 5's critique of Functionalism.

Frank Jackson's "The Knowledge Argument"

Few thought experiments have triggered more debate than the one developed by philosopher Frank Jackson (1943–) against Physicalism. This one shows, he claims, that there are facts about conscious mental experiences distinct from all the physical facts about the world.

Mary the Neuroscientist

Jackson states it this way:

> 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. She discovers, for example, just which wave-length 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" (Chalmers, 2002, p. 275).

In short, Mary knows everything physical about how the brain processes color information. Jackson then proceeds.

> What will happen when Mary is released from her black and white room or is given a colour television monitor? Will she *learn* anything or not? It seems just obvious that she will learn something about the world and our visual experience of it. But then it is 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 (Chalmers, 2002, p. 275).

Upon release, Jackson suggests, Mary will learn something: for example, what red looks like, or what it's like to see red. But she already knew everything physical here. So it follows that there's more to know here than the physical.

Jackson summarizes "The Knowledge Argument" as follows:

1. Mary pre-release knows everything physical there is to know about other people.
2. Mary pre-release does not know everything there is to know about other people.
3. Therefore, there are truths about people that escape the Physicalist story.

Three Points About "The Knowledge Argument"

1. The thought experiment can be generalized to all the other senses.
2. "The Knowledge Argument" shares much with Nagel's and McGinn's arguments.
3. "The Knowledge Argument" has generated a huge literature. We'll explore just a few of the most important Physicalist responses to it.

Three Physicalist Responses to "The Knowledge Argument"

1. One may deny that Mary actually gains any knowledge after all.
2. One may admit that she gains knowledge, but argue that this is only because she hadn't worked out all the implications of what she already did know.
3. One may admit that Mary gains new knowledge, but deny that it's the kind of "knowledge of new facts" that is relevant to the debate about Physicalism.

Deny Mary Gains Knowledge

Jackson's thought experiment admittedly pumps our intuitions the way he wants them to be pumped. But now, neuroscience is very far from "knowing everything physical there is to know" about color perception; and how can we be so sure that a future neuroscientist *would* actually learn something?

Admit That Mary Gains Knowledge

This next strategy grants that Mary will learn something upon her release, but only because she hadn't worked out all the implications of what she already knew. This argument denies that this means that Physicalism is false. Why? Because, it suggests, simply learning something doesn't mean that what you learn isn't already contained in what you already knew.

For example, most everyone knows that a triangle is a three-sided closed figure. But it may only be mathematicians who also know various mathematical truths about triangles. But those truths are already "contained" in the definition, even if you hadn't worked them out yet.

Similarly, perhaps Mary *wouldn't* know "what it's like to see red" merely on the basis of knowing everything about neural firings and the like. But that might only be because working out what it's like to see red on that basis might be very difficult—and not because the facts about the conscious experiences are not already contained in the physical facts, thus preserving Physicalism.

Admit That Mary Gains New Knowledge, But Not Knowledge of "New Facts"

Consider how Paul Churchland summarizes Jackson's argument:

1. Mary pre-release knows everything there is to know about brain states and their properties.
2. Mary does *not* know everything there is to know about sensations and their properties.
3. Therefore, by Leibniz's Law, sensations and their properties are not identical to brain states and their properties.

But now, Churchland objects, the type of knowledge involved in premise one is different from the type of knowledge involved in premise two. What makes premise one true is that Mary has what we might call "discursive" knowledge about the brain; but what makes premise two true is what we might call "knowledge by acquaintance" of the relevant brain states.

If so, then when Mary leaves the room, she merely gains a different kind of knowledge, or a different *way* of knowing, about the brain states she already knows about. But these need not be two different kinds of facts, one of which she knows before release and the other that she only learns after release; they are just two different ways of knowing the very same facts. And if she doesn't learn new facts on her release, then Physicalism remains untouched.

Conclusion

All four arguments of the past two lectures conclude that consciousness—qualia—does not fit into the purely physical world, and thus that Physicalism is false.

What they share is their reliance on what we can conceive or not; what we're capable of thinking about without invoking a contradiction. And so perhaps *all* of them ultimately rely on Kripke's identity principle: it's precisely because identities are necessary that the distinguishable concepts of the mental and physical undermine their identity.

But are such arguments *good* arguments?

Historically speaking, what we're capable of conceiving, and what we judge ourselves to be capable of conceiving, may well change over time. But still, we have to decide what we believe now, given our current conceptual capacities and the information available to us at this time. And so while these four arguments may not prove *conclusively* that the mental is distinct from the physical, they provide pretty good reason to accept that, from where we currently stand, the mental appears to be distinct from the physical.

FOR GREATER UNDERSTANDING

Questions

1. Can you devise a persuasive example of "contingent identity," contra Kripke's claim that all identities are necessary? For example, are there conceivable circumstances in which the Morning Star is not identical to the Evening Star, or heat is not molecular motion, or water is not H_2O? Would the Twin Earth thought experiment described in Lecture 10 be a case, for example, where water is not H_2O?
2. In what ways is Jackson's "Knowledge Argument" both similar to and different from Nagel's and McGinn's arguments, as presented in Lecture 11? Is it, in the end, more or less persuasive than those arguments, or equally so (or not so)? Why?

Suggested Reading

Jackson, Frank C. "Epiphenomenal Qualia" and "What Mary Didn't Know." *Philosophy of Mind: Classical and Contemporary Readings*. Ed. David J. Chalmers. Pp. 273–280. New York: Oxford University Press, USA, 2002.

Other Books of Interest

Kripke, Saul A. *Naming and Necessity*. Cambridge, MA: Harvard University Press, 1980 (1972).

Ludlow, Peter, Yujin Nagasawa, and Daniel Stoljar, eds. *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*. Cambridge, MA: MIT Press, 2004.

Goldberg, Sanford, and Andrew Pessin. Chapter 7: "Consciousness." *Gray Matters: An Introduction to the Philosophy of Mind*. Armonk, NY: M.E. Sharpe, 1997.

Websites of Interest

1. The Stanford Encyclopedia of Philosophy features an entry entitled "Qualia: The Knowledge Argument," which discusses the topic and provides opposing viewpoints. — http://plato.stanford.edu/entries/qualia-knowledge
2. The Saul Kripke Center was established at the Graduate Center of the City University of New York in September 2007 to promote the study of the intellectual achievements of Saul Kripke. The website includes lectures by Professor Kripke. — http://web.gc.cuny.edu/KripkeCenter

Lecture 13

Mind-Body Causation, Part II

The Suggested Reading for this lecture is Donald Davidson's "Mental Events" in *Philosophy of Mind: Classical and Contemporary Readings*, edited by David J. Chalmers, pp. 116–125.

Introduction

In Lecture 7 we looked at the problem of mind-body causation in its Cartesian Substance Dualism form, focusing on how seventeenth-century thinkers dealt with it. In this lecture we'll bring things up to the present day.

T.H. Huxley and Descartes's Mechanical Conception of Animals

English biologist Thomas Huxley (1825–1895) examined one of Descartes's more infamous doctrines: the idea that non-human animals are mere automatons, lacking consciousness. The starting point is the recognition that purely physical science can explain much about animal behavior. That raises the question of whether *everything* an animal does, it does purely because of its physical nature—without any role for mind or mentality.

Huxley noted that we cannot definitely prove that animals lack consciousness, but then again we can't prove definitively, either, that they do have consciousness, because consciousness is not the sort of thing whose presence can be directly known by third parties. But the overall weight of the evidence suggests that animals have some form of mentality.

Then why can animal behavior be explained *without* invoking their mentality? Huxley said:

> We believe [of lower animals], in short, that they are machines, one part of which (the nervous system) not only sets the rest in motion . . . but is provided with special apparatus, the function of which is the calling into existence of those states of consciousness which are termed sensations, emotions and ideas (Chalmers, 2002, p. 28).

So animals *are* machines, and their brains both run their bodies and produce "states of consciousness." But what are these states for, exactly, if their brains run their bodies? Huxley again:

> . . . [M]olecular changes in the brain are the causes of all the states of consciousness of brutes. Is there any evidence that these states of consciousness may, conversely, cause those molecular changes which give rise to muscular motion? I see no such evidence . . . The consciousness of brutes would appear to be related to the mechanism of their body simply as a collateral product of its working, and to be as completely without any power of modifying that working as the steam-whistle which accompanies the work of a locomotive engine is without influence upon its machinery (Chalmers, 2002, p. 29).

So, the physical brain is completely responsible for animals' behavior, and produces their mental states. But the mental states themselves *cause and do absolutely nothing*. They are mere by-products of the working of the brain—causally inefficacious, inert, powerless.

Human Beings and Epiphenomenalism

But Huxley doesn't stop there:

> It is quite true, to the best of my judgment, the argumentation which applies to brutes hold equally good of [human beings] (Chalmers, 2002, p. 30).

According to this "Epiphenomenalism," mental and physical states and events are equally real; but mental events are mere "epiphenomena," that is, events without any genuine causal powers of their own. Mental events are purely effects, and never causes, of *anything*.

Evaluation of Epiphenomenalism

Epiphenomenalism has some advantages, but it remains problematic.

First, it grants that the physical can cause the mental. But why isn't that as unbridgeable a gap as the reverse direction?

Second, however, something seems wrong with stripping the mental of all causal power altogether. Could the mental be real without having any causal powers? Further, if the mental has no causal power at all then it wouldn't make any difference if mental states existed or not; we'd thus lose all reason

to think that other human beings even have minds, and it would be a mystery why natural selection would support the development of mentality.

Donald Davidson and Two Theses About Mind-Body Causation

Let's now set aside Descartes's Substance Dualism in favor of Property Dualism. Our causation problem then will concern how mental properties and events enter into causal relations with physical properties and events. And *this* problem is made more difficult by two now widely accepted theses, stressed by American philosopher Donald Davidson (1917–2003).

1. Causation involves "laws" connecting causes and effects.
2. "The anomalism of the mental": There are no laws connecting the mental and physical.

If causation requires laws, and there are none between the mental and the physical, then it follows that there can be no causation between them either.

Thesis One

"Causation involves laws connecting causes and effects."

Why? Because the sheer fact that an individual event, X, is followed by another event, Y, doesn't entail that X caused Y; perhaps it was just random, or some other causal factor was present besides the X. It's only if *all* X's are always followed by Y's that we'd conclude that any individual X is a cause of some individual Y. But that's just to say that causation involves laws between general *types* of events.

Thesis Two

"The Anomalism of the Mental"

Davidson's argument here may be framed as a response to this question: Is it plausible to hold that all the people who have a particular specific intentional state also might share the same particular specific neural property?

Here's why Davidson says no.

Why You Should Accept the Anomalism of the Mental

How we ascribe intentional states to people differs from how we ascribe physical states, such as brain states, so as to rule out there being any laws connecting the physical to the mental.

The ascription of intentional states is deeply regulated, for example, by principles of rationality. So, we refrain from ascribing contradictory beliefs to people unless the evidence is such that we must. Indeed, in order to take someone even to be speaking meaningfully, we must take them to have consistent and coherent beliefs, without contradictions. Further, we also ascribe to people beliefs that follow logically from other beliefs we've ascribed to them, even when we lack further explicit evidence that they have these other beliefs.

Davidson's general point is that before we ascribe any given belief, we must consider the numerous other things the person believes. So the ascription of intentional states is a "holistic process": we don't and cannot have intentional states individually, one-by-one, in isolation from each other.

In contrast, ascribing physical or neural states to individuals is under no such constraints. We may ascribe neural states one-by-one, without any need to make these states "cohere" with each other.

Now suppose, however, that there *were* a law connecting the mental and the neural: a law that said, whenever this brain state occurs, then that intentional state occurs. If so, then whenever that brain state did occur, so would the intentional state—and then intentional states *could* be ascribed one-by-one. But that is precisely what we *cannot* do.

So there can be no laws connecting the mental and the neural.

A Paradox

Causation involves laws, and there are no laws connecting the mental and the physical. The conclusion seems inevitable: no causation between the mental and the physical.

Except that it's widely held that there *is* such causation.

Davidson next offers a way out.

Paradox Solution: Token-Identity

The earlier argument rules out, Davidson observes, laws that relate *types* of mental events with types of physical events. But we can preserve mental-physical causation if we look not at types of events but at individual "token" events.

First, note that the very same individual token event can be described in many different ways. We might describe some individual event both as "Fred's thinking that philosophy is fun" *and* as "Fred's neuron 17 firing rapidly."

Next, the very same event can "fall under" a law when described one way, but not when described in other ways. Suppose Hurricane Katrina destroyed my iPod; the event of "Andrew Pessin's iPod being destroyed" does not fall under any relevant laws of nature, but that very same token event, described in purely basic physical terms, may well.

So now, if you want to say that a mental event causes a brain event, then there must be a law connecting the two; but there are no mental-physical laws, there are only purely physical laws. So if the mental event causes a physical event, it must be by virtue of its *also being a physical event*.

That is, the event that can be described as "Fred's thinking that philosophy is fun" can also be described as "Fred's neuron 17 firing rapidly." And insofar as that event can be described physically, then it can fall under purely physical laws and cause other physical events. A mental event can cause a physical event, in short, only if the mental event is also, simultaneously, a physical

event—only if, in other words, there is a "*token* identity" between the mental and the physical.

"Anomalous Monism"

Davidson's position is known as "Anomalous Monism." It's "anomalous" because he denies there are mental-physical laws. And it's a "Monism" because it is opposing the "Dualism" we've spent much time discussing. All that really exists is one thing, the physical world—that's the "monism." But that one thing can also frequently be described in mental terms.

The New Mind-Body Causation Problem

Version 1: Anomalous Monism

To many philosophers, however, Davidson's theory doesn't solve the causation problem.

A mental event can cause a physical event if it is also itself a physical event. Your desire for water can cause you to get some water if that desire is identical to some token physical brain event. But now the fact that this event is *also* a mental event, a desire for water, seems to be irrelevant to its causing your move toward water. Davidson has reduced all causation to the relationship between physical events—leaving the mental features causally irrelevant.

So the new problem of mental-physical causation is this: can we get mental events or properties to count as causally relevant, *as* mental?

Version 2: Extrinsicness of the Mental

Recall the discussion of computers and the Computational Theory of Mind in Lecture 9. Computer causation is due to the physical properties of the circuits; what its signals might "mean" is irrelevant. Similarly, the Computational Theory of Mind treats the "syntax" of intentional states, the physical properties of the brain, as doing all the causing. But this too seems to leave the semantics, what the states are about, irrelevant. This neural state would cause the next neural state even *if* they had different meanings, or none at all. But then the fact that these states are also "intentional states"—that is, mental states—is irrelevant.

In general, the causal powers of a given state are "intrinsic" to it: the very nature of the electrical states, and the neural states, cause the relevant subsequent states. But their semantic aspects are "extrinsic" to them: that a given electrical state or neural state has a meaning is not built into the state itself. But if meaning, and therefore the mental, are "extrinsic" features of the physical states, then they are causally irrelevant.

Version 3: Causal Closure of the Physical

Most scientists consider the physical world to be causally closed: for any physical event, if you trace its causes and effects you'll always stay in the physical. But then the very nature of physical science rules out the mental as having any causal influence over the physical, *as* mental.

Conclusion

We've explored the contemporary dimension of the mind-body causation problem, focusing on Property Dualism—and leaning towards the conclusion that there seems to be no clear way for mental events to have genuine causal influence in the physical realm. If the mental is distinct from the physical, it seems, we must reject its causal efficacy and relevance. In the next lecture, we'll see how some thinkers find this a good reason to deny the reality of the mental altogether.

FOR GREATER UNDERSTANDING

Questions

1. Huxley thought that Epiphenomenalism applied equally well to animals and humans. But are there any reasons to think that humans are different from animals in this regard—that Epiphenomenalism might well be true of animals but not of us? Or is it more plausible to hold that it simply is true of neither?
2. Does Davidson's theory successfully resolve the "paradox" of reconciling mind-body causation with the anomalism of the mental and the law-like nature of causation? If not, then might one argue that those strange alternative theories of causation discussed in Lecture 7, due to Malebranche and Leibniz, become more attractive, at least to theists?

Suggested Reading

Davidson, Donald. "Mental Events." *Philosophy of Mind: Classical and Contemporary Readings*. Ed. David J. Chalmers. Pp. 116–125. New York: Oxford University Press, USA, 2002.

Other Books of Interest

Goldberg, Sanford, and Andrew Pessin. Chapter 4: "Mental Causation." *Gray Matters: An Introduction to the Philosophy of Mind.* Armonk, NY: M.E. Sharpe, 1997.

Huxley, Thomas H. "On the Hypothesis That Animals Are Automata, and Its History." 1874. *Philosophy of Mind: Classical and Contemporary Readings*. Ed. David J. Chalmers. Pp. 24–30. New York: Oxford University Press, USA, 2002.

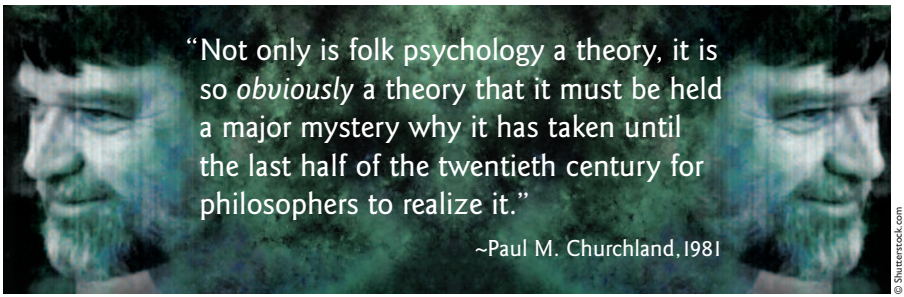
Websites of Interest

1. The *Stanford Encyclopedia of Philosophy* article on Donald Davidson by Jeff Malpas. — http://plato.stanford.edu/entries/davidson
2. The *Bancroft Library* at the University of California, Berkeley, provides a guide to the Donald Davidson papers through the *Online Archive of California*. — http://www.oac.cdlib.org/findaid/ark:/13030/kt9n39r8bg

Lecture 14

Eliminativism

The Suggested Reading for this lecture is Paul M. Churchland's "Eliminative Materialism and the Propositional Attitudes" in *Philosophy of Mind: Classical and Contemporary Readings*, edited by David J. Chalmers, pp. 568–580.



Introduction

Some philosophers argue that *there is no such thing as the mind*.

Paul Churchland and "Eliminative Materialism"

The failure of the Identity Theories suggests that identifying the mental with the physical is just not going to work. And that's because, as Churchland sees it, there's something wrong with our conception of mentality altogether: our basic mental categories are simply false:

> . . . [O]ur common-sense conception of [mental] phenomena constitutes a radically false theory, a theory so fundamentally defective that [it] will eventually be displaced, rather than smoothly reduced [or explained], by completed neuroscience (Chalmers, 2002, p. 568).

Our common-sense conception of the mind he calls "folk psychology." And the doctrine that folk psychology is false he calls "Eliminative Materialism."

Much of Churchland's argument is motivated by science. Many different scientific theories have posited the existence of various entities that subsequent theories came to reject. And so he claims that as brain science marches on, we'll realize that there simply are no such things as intentional states and qualia.

Three Reasons to Think That Folk Psychology Is False

1. Consider, first, the limits and failures of folk psychology. Churchland:

> [C]onsider the nature and dynamics of mental illness, the faculty of creative imagination, or the ground of intelligence differences between individuals. Consider our utter ignorance of the nature and psychological functions of sleep . . . Reflect on the common

> ability to catch an outfield fly ball on the run, or hit a moving car with a snowball. Consider the internal construction of a 3-D visual image from subtle differences in the 2-D array of stimulations in our respective retinas. Consider the rich variety of perceptual illusions . . . Or consider the miracle of memory, with its lightning capacity for relevant retrieval. On these and many other mental phenomena, [folk psychology] sheds negligible light (Chalmers, 2002, p. 571).

Folk psychology is thus sorely limited, at best a superficial account of our mental lives.

2. And now look at its history—which is, Churchland argues, one of "retreat, infertility, and decadence." Early humans inclined towards animism, the belief that nearly everything in the world had a mind; over time folk psychology became restricted to the higher animals only. But then even here it has simply stagnated! The folk psychology accepted by the ancient Greeks is the same as ours today, and we are no better at explaining human behavior than were they. Yet in the same time frame, science has made incredible strides in explaining the natural world.
3. Finally, folk psychology simply does not fit in well with the rest of science. We now know much about the nature of human beings, encompassing physics, chemistry, evolutionary theory, biology, and neuroscience. But folk psychology, Churchill writes,

> . . . is no part of this growing synthesis. Its intentional categories stand magnificently alone, without visible prospect of reduction to that larger corpus … Folk psychology's explanatory impotence and long stagnation inspire little faith that its categories will find themselves neatly reflected in the framework of neuroscience. On the contrary, one is reminded of how alchemy must have looked as elemental chemistry was taking form (Chalmers, 2002, p. 572).

Nor, as we've seen, do qualia fit any better than intentionality with our understanding of the brain.

Beyond Folk Psychology

Churchland next speculates about what our understanding of ourselves might be like once we do give up folk psychology.

On folk psychology, intentional states are closely related to language, as their contents nearly always seem expressible in sentences. One possibility is thus that the role of language in our mental life will simply be eliminated.

To illustrate, it's well-known that our brains have two hemispheres that differ in their basic functions. What's not as widely known is that the two hemispheres are closely linked, connected by fibers that let them communicate information at high speeds. In comparison, spoken English conveys information at only a fraction of that rate.

Well, if two hemispheres can communicate so efficiently, why couldn't two distinct brains do it as well? Maybe one day people will wear brain caps that transmit information to others' brains, and vice versa. And how exactly will such people understand and conceive of other individuals? Churchland answers:

> In roughly the same fashion that your right hemisphere "understands" and "conceives of" your left hemisphere—intimately and efficiently, but not [linguistically]! (Chalmers, 2002, p. 579).

So one possible outcome of the advance of science: ordinary language falls away, and with it the whole notion of linguistically dependent intentional states.

Yes, But . . . Qualia?

But what could be more obvious than the reality of *qualia*?

Two Ways of Denying the Existence of Qualia

What motivates the "elimination" of qualia is Physicalism. But we may distinguish two different Physicalist positions.

1. "Scientific Qualia Eliminativism." This more modest version merely claims that qualia have no place in the science of psychology.

   Why? First, qualia are subjective, and no one has access to anyone else's qualia; so qualia can't directly be studied by scientists. But second, qualia are simply not relevant to the explanation and prediction of our behavior, for they make no testable *difference* to that behavior. And so scientific psychology—which aims to explain and predict—has no place for them.

2. "Philosophical Qualia Eliminativism." This stronger version claims there simply are no qualia, period.

   The general strategy here is, first, to clarify exactly what properties qualia are alleged to have, then to argue that nothing actually has those properties.

Daniel Dennett on Quining Qualia

According to "The Philosophical Lexicon," a satirical dictionary of philosophical eponyms, "to quine" (in honor of Harvard philosopher Willard Quine, 1908–2000) means "to deny resolutely the existence or importance of something real or significant." And "quining qualia," Dennett insists, is the only way for a Physicalist to go once we give up identifying qualia with anything physical.

Task One: Sketching the Properties of Qualia

Qualia are generally thought to have various properties.

1. They are "directly" or "immediately" graspable during our conscious experiences.
2. We are infallible about them.
3. They are "private": no one can directly access anyone else's qualia.

4. They are ineffable.

5. They are "intrinsic" and "simple" or "unanalyzable."

Dennett next argues that nothing quite fits that set of descriptions.

Task Two: Nothing Has Those Properties

To do this, Dennett devises a series of "intuition pumps"—thought experiments designed to explore our intuitions about conscious experience.

So, for example, he introduces the inverted qualia case: two people could have different qualia and yet all the same external physical behavior. But now the qualia supporter might then present an "intrapersonal" variation. Suppose a devious neurosurgeon fiddles with your brain and you wake up to discover that the grass looks red. Wouldn't this be a case where we could confirm the reality of qualia—by noticing how the qualia have changed while every other aspect of our conscious experience remains the same?

Not quite, Dennett replies via the next intuition pump, "alternative neurosurgery." In fact there are two different ways the neurosurgeon might have accomplished the inversion above.

First, she might have tinkered with something "early on," so that the signals coming from the eye when you look at grass contain the information "red" rather than "green." This would result in a genuine qualia inversion.

But, alternatively, she might have instead tinkered with your memory. Here your qualia would remain the same, but your memory would be altered so that your current green experience would contradict your earlier memories of grass. Note, you would still feel that "the color of grass has changed"; only here it isn't the qualia that has changed, but your memories.

But now, would you be able to tell which of these scenarios is correct? No: your perceptual experience tells you that *something* has changed but not whether your *qualia* have changed. But this leads to an important conclusion. Dennett:

> Since [by hypothesis] the two different surgical invasions can produce exactly the same introspective effects while only one operation inverts the qualia, nothing in the subject's experience can favor one of the

> hypotheses over the other. So unless he seeks outside help, the state of his own qualia must be as unknowable to him as the state of anyone else's qualia. Hardly the privileged access or immediate acquaintance or direct apprehension the friends of qualia had supposed [qualia] to enjoy! (Chalmers, 2002, p. 231).

It's questionable, in short, that we have direct, infallible access to our conscious experience.

The Experienced Beer Drinker

Many people think of beer as an acquired taste: one's first sip is often unpleasant, but one gradually comes to enjoy it. But wait, Dennett asks—what's the "it" here? Compare the flavor of that first taste with the flavor now. Does the beer taste exactly the same both then and now, only now you like that taste whereas before you disliked that very same taste? Or is it that the way beer tastes gradually shifts—so that the taste you did *not* like at the beginning is *not* the very same taste you now like at the end?

In fact most people simply cannot *tell* which is the correct analysis. But that is to give up again on the idea that we have special and infallible access to our qualia.

Further, when forced to choose, many people feel that the second analysis is more plausible. But then if one's *reactions* to an experience are in any way constitutive of it, the experience is not so "intrinsic" after all—and another qualia property falls.

The Inverted Goggles

This one is grounded in a real-life phenomenon.

Scientists have devised special eyeglasses that invert up and down for the wearer. When you put them on, everything looks upside down. When subjects first don them they can barely walk around without stumbling.

But when subjects wear these goggles for a while, something surprising occurs. They adapt, and become able to walk around as easily as before. When you ask them whether they adapted by re-inverting their visual field or whether they simply got used to walking around in an upside-down world, *they can't say.*

So as in our beer-drinking case, either we simply do not have the special, infallible access to our qualia that would allow us to distinguish the two cases, or, perhaps, the way the world looks to us is actually a function of how we respond to the world—in which case qualia are not "intrinsic" properties of experience.

Dennett's Conclusion

Dennett's intuition pumps chip away at all the alleged properties of qualia, suggesting that, in the end, nothing in fact fits the description of qualia with which we began.

So qualia don't exist.

Conclusion to the Course

In this course we've looked at the nature of the mind and the mental, and the relationship between them and the physical. We examined intentionality, and qualia, and whether such states or events might be identifiable with anything physical. Along the way we discussed causation, and computers, and zombies. If there was a theme throughout, it's that the question of how exactly the mental fits into the physical world is far from being solved. This led to our final lecture—in which we explored the idea of simply denying the existence of the mental altogether.

So where does one go from here?

We're really just at the dawn of the sciences of the mind, and these developing sciences will, over the next decades, reveal countless fascinating facts about our minds that no one could possibly have anticipated. And when they do so, the philosophers will be there trying to figure out, as is their wont, what it all means.

FOR GREATER UNDERSTANDING

Questions

1. Consider this response to Churchland: "Folk psychology couldn't be false—for if it were, there would be no such thing as 'beliefs.' But then it wouldn't be true that you believe that folk psychology is false! So it's logically impossible to deny the truth of folk psychology." Is this a persuasive objection to Churchland's critique of folk psychology?
2. Do Dennett's "intuition pumps" succeed in challenging some of those alleged properties of qualia? If not, why not? If so, could it be possible to argue that qualia still do exist, even if, perhaps, they don't have all the properties one might originally have thought they had? Or are those properties really *essential* to the existence of qualia?

Suggested Reading

Churchland, Paul M. "Eliminative Materialism and the Propositional Attitudes." 1981. *Philosophy of Mind: Classical and Contemporary Readings*. Ed. David J. Chalmers. Pp. 568–580. New York: Oxford University Press, USA, 2002.

Other Books of Interest

Dennett, Daniel C. "Quining Qualia." 1988. *Philosophy of Mind: Classical and Contemporary Readings*. Ed. David J. Chalmers. Pp. 226–246. New York: Oxford University Press, USA, 2002.

Goldberg, Sanford, and Andrew Pessin. Chapter 2: "The Mind-Body Problem." *Gray Matters: An Introduction to the Philosophy of Mind*. Armonk, NY: M.E. Sharpe, 1997.

Websites of Interest

1. The *Science Network* website provides a video interview with the husband-and-wife team of neurophilosophers Paul and Pat Churchland entitled "From the Engine of Reason to the Seat of the Soul: A Brain-Wise Conversation with Neurophilosophers Pat and Paul Churchland." — http://thesciencenetwork.org/programs/the-science-studio/from-the-engine-of-reason-to-the-seat-of-the-soul-a-brain-wise-conversation
2. The Open Directory Project is a comprehensive human-edited directory of the Web. This section provides a list with links to websites focused on philosophy of mind. — http://www.dmoz.org/Society/Philosophy/Philosophy_of_Mind

COURSE MATERIALS

Suggested Readings

Chalmers, David J., ed. *Philosophy of Mind: Classical and Contemporary Readings*. New York: Oxford University Press, USA, 2002.

Fodor, Jerry A. *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press, 1992.

Ryle, Gilbert. *The Concept of Mind*. 60th anniv. ed. New York: Routledge, 2009 (1949).

Stich, Stephen P., and Ted A. Warfield. *The Blackwell Guide to Philosophy of Mind*. Oxford: Wiley-Blackwell Publishing, 2003.

Other Books of Interest

Ariew, Roger, and Eric Watkins, eds. *Modern Philosophy: An Anthology of Primary Sources*. 2nd ed. Trans. Roger Ariew and D. Garber. Indianapolis, IN: Hackett Publishing Company, 2009.

Dahlbom, Bo. *Dennett and His Critics: Demystifying Mind*. Malden, MA: Blackwell Publishers, 1995.

Dennett, Daniel C. *Consciousness Explained*. Boston, MA: Back Bay Books, 1992.

Fodor, Jerry A. *LOT 2: The Language of Thought Revisited*. Oxford, UK: Oxford University Press, 2008.

Goldberg, Sanford, and Andrew Pessin. *Gray Matters: An Introduction to the Philosophy of Mind*. Armonk, NY: M.E. Sharpe, 1997.

Kripke, Saul A. *Naming and Necessity*. Cambridge, MA: Harvard University Press, 1980 (1972).

Loewer, Barry, and Georges Rey, eds. *Fodor and His Critics: Meaning in Mind*. Malden, MA: Blackwell Publishers, 1992.

Ludlow, Peter, Yujin Nagasawa, and Daniel Stoljar, eds. *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*. Cambridge, MA: MIT Press, 2004.

Pessin, Andrew, and Sanford Goldberg, eds. *The Twin Earth Chronicles: Twenty Years of Reflection on Hilary Putnam's "The Meaning of Meaning."* Armonk, NY: M.E. Sharpe, 1996.

Ross, Don, Andrew Brook, and David Thompson, eds. *Dennett's Philosophy: A Comprehensive Assessment*. Cambridge, MA: MIT Press, 2000.

Skinner, B.F. *Walden Two*. New rev. ed. Indianapolis, IN: Hackett Publishing Co., Inc., 2005 (1948).

Smythies, John R. *The Case for Dualism*. Ed. John Beloff. Charlottesville, VA: The University of Virginia Press, 1989.

These books are available online through www.modernscholar.com or by calling Recorded Books at 1-800-636-3399.

RECORDED BOOKS

The study of philosophy is among the most popular course topics in colleges and universities around the world. The Modern Scholar also offers the following courses on philosophy and human thought.

Plato and Aristotle: The Genesis of Western Thought

Professor Aryeh Kosman—Haverford College

If one is to truly understand the birth of Western thought, the works of Plato and Aristotle must be the starting point. To understand who we are as human beings, we must begin to understand the work and concepts articulated by these two early thinkers. For it is in their writings, more than anywhere else, that we begin to see the development of the Western cast of mind. In this course, we'll investigate the essential texts of these two early philosophers of Ancient Greece. Far from simply making a list of their ideas, we'll focus on thinking through the premises they put forth. Our goal will be to arrive at a solid understanding of the philosophy of these two men and how their legacies continue to influence us today.

Ethics: A History of Moral Thought

Professor Peter Kreeft—Boston College

What is good? What is bad? Why is justice important? Most human beings have the faculty to discern between right and wrong, good and bad behavior, and to make judgments over what is just and what is unjust. But why are ethics important to us? This course looks at our history as ethical beings. We'll travel into the very heart of mankind's greatest philosophical dilemmas—to the origins of our moral values and the problem of ethics. Are ethics universal, absolute and unchanging—or are they culturally relative, changing, and man-made? Furthermore, we'll delve into the creation of ethical systems—not just for ourselves, but also for society at large.

Discovering the Philosopher in You: The Big Questions in Philosophy

Professor Colin McGinn—Rutgers University

Of all the branches of intellectual enquiry, many find philosophy to be the most esoteric and difficult to grasp. Yet, the basic questions of philosophy, from logic to ethics, from the human mind to God, have been pondered by people around the world for centuries. For surely everyone, on some level, wants to know what the ultimate nature of the world is, where ethical truth comes from, what the meaning of life is, and whether or not we can really know anything. In this course we will explore these main philosophical problems, with each lecture investigating different questions. Our goal, in the end, is to uncover and exercise the philosopher that lies within each of us.

Creating Humans: Ethical Questions Where Reproduction and Science Collide

Professor Alexander McCall Smith—The University of Edinburgh

Human reproduction is about life and its perpetuation. If there is anything that we have to take seriously from a moral point of view, then surely it is human life. We value life because it is all that we have which is our own. We construct elaborate systems of belief about it; we guard it through rules we have devised for it; it is everything to us. How we begin human life, how we bring it into existence is a matter of the most profound importance. In this course, we will discuss the various moral aspects of human reproduction from methods of conception to methods of ending a pregnancy. We will discuss the moral, cultural, legal, and political influences on reproduction as well as the scientific advances in reproductive technology.

Ideas That Shaped Mankind: A Concise History of Human Thought

Professor Felipe Fernández-Armesto—Oxford University

Ideas That Shaped Mankind, from internationally respected historian Felipe Fernández-Armesto, explores the notion that man's capacity to produce ideas brings about sweeping changes in the world. This ability, seen most profoundly in individual moments of genius—or equally startling moments of chance—is what separates humans from the animals and allows humans to re-imagine the world in ever more complex designs. From the earliest ideas, including cannibalism and the idea of farming, to theories of relativity and chaos, ideas reshape the world in surprising and unexpected ways. Science, agriculture, religion, art, politics—Professor Fernández-Armesto examines all these areas of thought and the moments in time when man's fertile intellect produced the sparks that set off blazes of change that would forever alter the course of human history.

Philosophy and the Law: How Judges Reason

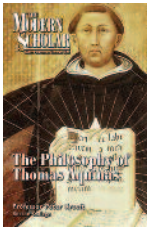
Professor Stephen Mathis—Wheaton College

Do judges deduce their decisions from legal rules and principles, or do they decide cases based on what is fair given the facts at hand? The latter view, held by Legal Realists, serves as the starting point for Professor Stephen Mathis' eye-opening look at how judges reason. In this compelling lecture series, the esteemed professor addresses such issues as whether the law is distinct from morality. Professor Mathis also attempts to identify a view that offers guidance to judges in deciding cases, and one that will provide the tools people need to evaluate the interpretations and decisions judges make.

Visions of Utopia: Philosophy and the Perfect Society

Professor Fred E. Baumann—Kenyon College

Since ancient times, philosophers have struggled with the concept of the ideal society, or utopia. Many have contributed to the widely varying possibilities for just what such a system might entail. Religious, economic, and political structures all help to shape the composition of these utopias, and as these visions are shared, they impact the way subsequent utopias are envisioned. In this fascinating series of lectures, Professor Fred E. Baumann explores the "perfect" societies that have been described by great thinkers throughout history. His analysis provides insight on the nature of utopias and their place in society.



The Philosophy of Thomas Aquinas

Professor Peter Kreeft—Boston College

An enthusiastic admirer of the philosophy of Thomas Aquinas, professor and philosopher Peter Kreeft brings the full measure of his skill to these lectures. By providing examples of present-day situations as well as historical references, Professor Kreeft details the rational thought and precise literary talent that established Aquinas as the foremost thinker of his time. Aquinas's *Summa Theologica* has given theologians and philosophers much to discuss since the thirteenth century. Peter Kreeft explains why.

These courses are available online through www.modernscholar.com or by calling Recorded Books at 1-800-636-3399.

A Recorded Books Production
www.modernscholar.com • 1-800-636-3399