Introduction to Philosophy

INTRODUCTION TO PHILOSOPHY

MATTHEW VAN CLEAVE, PAUL JURCZAK, CHRISTOPHER SCHNECK, AND DOUGLAS SJOQUIST

CONTENTS

Chapter 1: External world skepticism

Chapter 2: The problem of other minds

Chapter 3: The mind-body problem

Chapter 4: Personal identity

Chapter 5: The problem of free will and determinism

Chapter 6: Philosophy of religion

Chapter 7: The ethics of belief

Chapter 8: Human well-being

Chapter 9: Post-modernism

Chapter 10: Property

Chapter 11: Capitalism, Communism, & Bernie Sanders

Chapter 12: Buddhism

Chapter 13: Aesthetics

PREFACE

Matthew Van Cleave

This textbook grew out of an introduction to philosophy course that we designed, initially the result of the urging of our former associate dean, Eric Snider, who was tragically killed while riding his bike to work in late July, 2016. Eric never got to see the project he initiated come to fruition, but we greatly benefitted from his support while he was our associate dean. It is not every day that philosophers are lucky enough to have an administrator who was himself an excellent philosopher. For the years he was in our department at Lansing Community College (LCC), Paul Jurczak and I benefitted from his special concern for the humanities, which was a bright spot in an otherwise gloomy environment that regularly fails to understand the value and purpose of the humanities. Eric's specialization was ancient Greek philosophy. He wrote a dissertation at Vanderbilt in the 1980s with Alasdair MacIntyre on Aristotle's account of practical reasoning. Undoubtedly this textbook would have been much richer if it contained his input. We dedicate this textbook to Eric Snider (1957-2016).

This textbook also was supported by a grant that was secured by Regina Gong when she ran the Open Educational Resources (OER) grant program at LCC. Although she has moved onto bigger and better things at Michigan State University, where she is the OER and Student Success Librarian, we are grateful for her tireless work in support OER during her tenure at LCC. Especially at a community college, where a higher proportion of students are first-generation college students or from less privileged backgrounds, providing high-quality, low-cost textbooks is essential for helping to secure equity in higher education. We are grateful for Regina's vision on this and for getting us all on board.

Matthew Van Cleave

August 18, 2019

INTRODUCTION

Matthew Van Cleave

Philosophy is many things to many people and so writing a general introduction to philosophy textbook is bound to itself be an exercise in meta-philosophy. Because this is so, there will be disagreements not only about what is in the textbook, but also about what it leaves out. In the process of editing the chapters for this text, the philosophers in my department (Paul Jurczak, Christopher Schneck, and Leanne Kent) had occasion to hash out some of these disagreements. I believe the resulting textbook is better for it, not to mention that we as philosophers are all better for it. As will perhaps be obvious for those who read the textbook, the different chapters have very different styles. We consider this diversity a virtue of the textbook, although we understand that it does make it bit more unwieldy. For example, Paul and I have very different backgrounds in philosophy. Paul reads French very well and knows a lot about Merleau-Ponty; I read no French, but am fairly engaged with the cognitive sciences. Paul knows a lot about history, literature, and film; I don't, but can tell you more than you'd like to know about what's wrong with Jerry Fodor's asymmetric dependence account of mental representation or why Saul Kripke's causal account reference is important. As philosophers may infer based on these descriptions, I was schooled within the "analytic" tradition of philosophy, whereas Paul's interests lie more within the "Continental" tradition of philosophy. These differences speak to the diversity of what philosophy is and they are reflected in this textbook.

Organizationally, this textbook follows a topic-oriented, rather than historical, approach. Of course, some of the chapters do delve into the history of philosophy—in particular, Paul's chapters. What I have tried to do in the chapters I've written for the textbook is to stick to the things that have worked really well for me over the years in teaching. By "worked well" I mean that the topics, examples, and orienting questions have gotten students involved and active in the course. I have found that I can best accomplish this by being very specific, rather than speaking more generally about a philosophical issue. To that end, much of the content of my chapters concerns rich examples and detailed arguments from specific philosophers. To me, a good textbook for a philosophy class is one that creates fruitful conversation—one that sets up good Socratic dialogue. That is why Plato's dialogues typically work so well: they don't give you the answers, but rather set up problems for further discussion. This is one of the things that students often find difficult about a philosophy course but that is simultaneously vital for introducing students to philosophy: intellectual discussion about perennial conceptual disputes. Students wanting to "know the correct answer"

have yet to really understand what philosophy is about. My goal in my courses, and what I have tried to do in this textbook, is to facilitate that better understanding of what philosophy is—not just in terms of the issues that are debated by philosophers, but also in terms of the tools that philosophers use to do it. First

and foremost, the use of rational argumentation, counterexamples, and thought experiments. Philosophers will recognize this as a distinctly "analytic" approach to philosophy, but my colleagues' chapters are no less interested in rational argumentation, even if they don't always go about it as explicitly as I do in my chapters. One technique that I regularly use in class (and have also used in this textbook) is reconstructing arguments with numbered statements, as one would do in a logic class. Again, this reflects my particular analytic approach to teaching philosophy and other chapters do not use this apparatus, but one could also reconstruct arguments in those chapters, if that was something one was inclined to do.

This open textbook is a work in progress. There are chapters that are still being drafted and others that we would yet like to draft, including chapters on: Kantian versus utilitarian normative ethical theories, feminist philosophy, philosophy of race, philosophy of science, epistemic injustice, philosophy of language, the epistemology of testimony, metaethics, and justice. It is also a collaborative effort. So far, there are six philosophers that have been involved in either drafting or editing the chapters. If you are a philosopher who would like to contribute any of the above chapters (or any others) to this open textbook, I encourage you to send me an email at vancleam@lcc.edu to discuss that prospect further.

CHAPTER 1: EXTERNAL WORLD SKEPTICISM

THE PROBLEM OF EXTERNAL WORLD SKEPTICISM

Matthew Van Cleave

In the Wachowskis' 1999 film, *The Matrix*, Morpheus asks Neo:

> "Have you ever had a dream, Neo, that you were so sure was real? What if you were unable to wake from that dream? How would you know the difference between the dream world and the real world?"

Philosophers would recognize that this question is not a new one. In fact, René Descartes (1596-1650, pronounced day-CART) asked this exact question 350 years before the Wachowskis in his *Meditations on First Philosophy*, published in 1641. The question, it turns out, is a philosophically deep one and the problem is that it seems that there is no good way to tell whether one is in fact dreaming or not.

You might think that the answer to the question of whether I am dreaming right now is an easy one: of course I'm not dreaming because I remember waking up this morning and eating breakfast! Or: I know I'm not dreaming because I can pinch myself and it hurts! Or: I am reading this philosophy textbook right now and that is not something that I would ever do in my dreams! However, all of these responses seem to miss the deeper challenge of the I-am-dreaming-right- now scenario. For all you know, your whole life has been one long, vivid dream that you have never awoken from. Sure, you have gone to sleep and woken up and even had dreams when you've "slept," but this has all happened within one long meta-dream, which has been your life to date. All kinds of questions could be raised about such a possibility, such as: if I have been dreaming my whole life, how could I eat? And if I haven't eaten how could I survive to continue dreaming?

These are good questions, but the modern skeptic has an easy solution to them, since there are many other skeptical scenarios that don't raise these questions.

The scenario presented in *The Matrix* is a good example, so let's switch to it. Imagine that your whole life has been a kind of computer simulation—your conscious mind interacts not with the "real world" but with a computer simulation. Within the computer simulation are all kinds of objects (trees, houses, cars, computers, other people, etc.) that appear to you exactly as they would in the "real world"—i.e., in a non-simulation world—the kind of world that

we think that we *do* live in. I can imagine someone asking: wouldn't I be able to tell that a simulated coffee cup wasn't a real coffee cup? Wouldn't I be able to tell that it is a simulation? Consider what Morpheus says to Neo in *The Matrix*: "What is real? How do you define 'real'? If you're talking about what you can feel, what

you can smell, what you can taste and see, then 'real' is simply electrical signals interpreted by your brain." The point is that the object itself is not actually necessary in order for us to have a perception of it. If the coffee cup is just a pattern of brain activity, then we should be able to stimulate the brain in that particular way and thus produce a perception of a coffee cup—even if there is no coffee cup. Hallucinations and illusions are a kind of everyday case in point. **External world skepticism** is the claim that I cannot know whether I live in a simulation or a non-simulation because from inside my own conscious experience I would not be able to distinguish between the two different scenarios. After all, I would have exactly the same experience in the simulation world as I would in the non-simulation world. If this is so, then it seems that I cannot know if there exists anything outside of my own experience and perceptions. If I am really just in a simulation (i.e., a kind of dream world) then it seems that the objects with which I am interacting are not real objects—the dog bite in my dream is not the same as the real life dog bite. Of course, the objects in the simulation would *seem* real to me but in some deep sense they wouldn't be real. Simulated dog bites are not real dog bites.

For example, if I had a dream that I was drinking coffee out of my favorite coffee mug and if I said to myself in that dream, "I am drinking coffee right now," then what I have said would seem to be false (assuming I didn't know that I was dreaming). Likewise, if in my dream my friend Bob kicked me and I formed the belief, "Bob kicked me" I would seem to have a false belief. Bob didn't really kick me; I just dreamed he did. And I am not really drinking coffee; I just dreamed I did. In general, dreaming that x happened is not the same as x actually happening. So if in my vivid dream I come to believe any of these things, then those beliefs are false. In fact, all of my beliefs would seem to be false since all of the things I'm dreaming about are not real. The same would seem to be true if I am living in a simulation since the objects in that simulation are not "real" in the way we think that non-simulation objects are real.

Another way of laying out the problem of external world skepticism is to note that our experience does not allow us to distinguish the nature of things outside of us (since there's no way for us to, so to speak, get outside of our own minds or experience) and that, since there are multiple ways that our experiences could

be caused (dreams, hallucinations, computer simulations, etc.), we cannot know which of those things is the cause of our experiences. But since this is so, it poses a deep problem for our knowledge. After all, everyone thinks that they can know that we live in a world of normal, physical objects and that these objects are what we are experiencing—i.e., they are what are causing our perceptions. But as soon as we become aware of the radically different scenarios that could equally explain our experience—what we can call **skeptical scenarios**—we must acknowledge that we don't know which of these scenarios we are actually in. If I cannot know whether I am a person with a body and who interacts with real physical objects or a brain in a vat that is hooked up to a supercomputer which stimulates my brain to produce perceptions of what I think is the real physical world (but actually isn't), then I cannot know whether there is an external world or not. That is the problem of external world skepticism.

We can put the problem of external world skepticism into a **reconstructed argument** like this:

1. I know there is an external world only if my experience allows me to distinguish between the real world and skeptical scenarios.
2. My experience does not allow me to distinguish between the real world and skeptical scenarios.
3. Therefore, I cannot know for certain that the external world exists.

The logic of this argument is airtight. That is, if we accept the truth of the premises, then the conclusion has to be true. The above argument has the following sort of logical form:

1. If A is true then B must be true
2. B is not true
3. Therefore, A is not true

As you should be able to see, there's nothing wrong with the logic of this argument. But that leaves open whether or not the premises are true. So are the premises true? Premise 1 seems to be true. After all, if I am merely dreaming I am in my friend Kali's living room right now, then I am not really in her living room and thus can't know that I am.[1] Premise 2 also seems to be true: if my experience is confined to my own head, it seems that I am not able to see what is "really" outside of me. One might say at this point that I can compare my experience to the experience of others and in this way I could determine whether or not there are objects outside of my experience. Certainly, if I am experiencing a pink unicorn and I ask everyone else around me whether they see the pink unicorn and everyone looks at me like I'm crazy, then I have good reason to doubt whether the pink unicorn is real. Probably it is an illusion or hallucination. But this response neglects the fact that *if* I am in a dream/simulation, then my experience of you is itself part of the dream/simulation. Since you are part of the dream/simulation, our agreement concerning our perceptions doesn't prove that the objects we seem to be perceiving are real, since we would both be under the same illusion.

Notice what the problem of external world skepticism is *not*. The external world skeptic is not claiming that we actually *are* living in a computer simulation, that we actually *are* dreaming, or that we actually *are* a brain in a vat. Rather, they are claiming that we cannot *know* that we *aren't* in such a skeptical scenario.

And if we cannot know such a fundamental thing as this, then we cannot really have any knowledge about the world outside of us. None of this is to deny that we have experiences. The problem of external world skepticism does not question our experiences themselves. Rather, it questions whether we can know what those are experiences *of*—whether we are experiencing what we think we are (a normal, physical world). The

---

1. Note that that fact that I'm dreaming does not imply that I'm *not* in Kali's living room. For example, someone might have kidnapped me while I was sleeping (and having this dream) and put me in Kali's living room. It would still seem to follow, however, that I didn't really *know* I was in Kali's room, since my true belief would have been a kind of fortuitous accident.

skeptic's claim isn't that we are in a skeptical scenario but that *we cannot know we're not*. And that is supposed to undermine all (or almost all) of our knowledge about the world.

Study questions

1. True or false: The external world skeptic claims that we are probably living in a computer simulation or dreaming.
2. True or false: External world skepticism claims that there is no external world; the only reality is our own perceptions.
3. True or false: External world skepticism claims that we cannot *know* whether there is an external world (similar to what we perceive) that exists outside our perceptions.
4. True of false: According to external world skepticism, if our experience does not allow us to distinguish skeptical scenarios from (what we think of as) the real world, then we cannot know there is an external world.
5. True or false: It is possible that the argument for external world skepticism has flawed *logic*.

For deeper thought

1. Suppose that you were dreaming that you were eating ramen noodles; would it follow that you were not really eating ramen noodles? Why or why not?
2. If you were dreaming (and believed in your dream) that you were in Pensacola, FL and it turned out that you really were in Pensacola, would that mean you knew (in your dream) that you were in Pensacola? Why or why not?
3. How would you respond to the external world skeptic? Can we in fact show that there exists outside of our minds a world that is roughly similar to what our perceptions tell us? Why or why not?

Responses to external world skepticism

It seems difficult to refute the external world skeptic. That is, it seems difficult to prove that there are objects that exist independent of our perception since we cannot get "outside" our perception to observe them. Nonetheless, philosophers over the centuries have attempted to refute the external world skeptic. We will consider three different answers to the skeptic: one from G.E. Moore in the early 20th century, one from Ludwig Wittgenstein, also from the early 20th century, and one from David Chalmers in 1999.

Before we look that those responses, let's consider one famous argument that there are some things we can

know even if we are in a skeptical scenario. René Descartes has a famous five-word that you may be familiar with: I think, therefore I am.[2] Here is Descartes's point: even if I am dreaming, and thus wrong that there is a physical world outside of me, I cannot be wrong that I exist as a thinking thing. Why not? Well assume that I am dreaming. In that case, there must be something that does the dreaming since if there are dreams then there must be a dreamer. Descartes's point is that I cannot be mistaken about the fact that I have perception and thoughts *when I am having those perceptions and thoughts*. Although I may be mistaken (if I'm dreaming or in some skeptical scenario) about all kinds of things *outside* of my experience, as long as I'm focusing on the experiences themselves, I cannot be mistaken that I am having them. If something is having those experiences, then that thing—my mind— most certainly exists. My mind may not have an attached body and may not look anything like what I think "I" look like, but I cannot be wrong that there is a thing that has experiences.

Consider the difference between the following two statements:

- I am eating a juicy steak
- I seem to be eating a juicy steak

Descartes would say that although you cannot be certain that the first one is true (since you could be merely dreaming that you are, or in a simulation, etc.), you can be certain that the second is true since it is about *your experience* rather than the objects outside of you. Even if you are merely dreaming that you are eating a steak, your statement that you seem to be doing so is true, whether you are dreaming or not. The point Descartes is trying to make is that I can be

certain that my mind exists even if I cannot be certain that anything exists outside of my mind. What I am unsure about is *not* whether my mind exists, but whether there is an external world. There cannot be any doubt, Descartes thinks, that there is an internal world—i.e., a mind that is doing the thinking and perceiving. This is the meaning of the famous dictum, *I think, therefore I am*.

Descartes's argument that I cannot be mistaken that I exist (as a thinking thing) doesn't provide a solution to external world skepticism, it only reinforces the problem. The external world skeptic is not skeptical about the existence of your thoughts and perceptions. Rather, she is skeptical about whether the external world exists. The skeptic's point is that you cannot *know* that there is such a world outside your mind.

---

2. Actually, it is only a three-word argument in Latin (the language in which Descartes—and all medieval philosophers before him—wrote): *cogito ergo sum*. Descartes' argument is sometimes referred to the "cogito argument," "cogito" being Latin for "I think."

Moore's response to external world skepticism

G.E. Moore was a 20th century British philosopher who thought he could prove that there was an external world. His argument is simple, but it will take a little bit of unpacking. Here is Moore's argument (you have to imagine Moore demonstrating it to you with gestures as he says it):

1. Here is a hand [Moore holds up one of his hands]
2. Here is another hand [Moore holds up his other hand]
3. Therefore, two human hands exist at this moment (from 1-2)
4. This same argument could be made for any object I could hold in front of you.
5. Therefore, external objects exist (from 3-4)

If you've followed the problem of external world skepticism, there's something that seems fishy about Moore's argument. Moore's argument seems to commit the fallacy of **begging the question**. The problem is that premises 1 and 2 already seems to assume that the objects he/we are perceiving exist outside our minds in roughly the way we perceive them (what else could justify the inference to line 3)? Alternately, if premises 1 and 2 are simply referring to our perceptions, then line 3 doesn't seem to follow since "hands" are the kinds of things which we conceptualize as existing independently of our perceptions of them (i.e., my hands still exist even if I'm sleeping or not currently perceiving them). But the skeptic would protest that Moore is not entitled to make any of these assertions because these assumptions are precisely what the skeptic is trying to put into question! This is the informal logical fallacy of begging the question—to assume as one of your premises, the conclusion that your argument attempts to establish.

Although Moore doesn't think he is begging the question, it is important to understand the general strategy that Moore uses to defend his response to external world skepticism. That response turns on a comparison of our level of subjective certainty about certain propositions rather than trying to show that we *aren't* in a skeptical scenario. In fact, Moore accepts that we cannot prove that we aren't in a skeptical scenario, but he doesn't think that we need to prove that in order to know that there is an external world. Which of the following are you more certain of, Moore asks:

- Here (holding up your own hand) is one of my hands
- It is possible that I am actually in a computer simulation right now
- If I am in a computer simulation, then I do not know that I have hands

Moore's point is that he is more certain of the truth of the first proposition than he is of the second or third (both of which are premises in the argument for external world skepticism). Moore thinks that because he is more certain of the premises of his argument than he is of the premises of the skeptic's argument, his argument is therefore a better argument. Moore's response to external world skepticism is thus what I would call an

indirect response because it doesn't attempt to refute the skeptic's argument but, rather, attempts to put a competing argument forward. The conclusion of Moore's argument is that we can know that there is an external world; the conclusion of the argument for external world skepticism is that we *can't* know that there is an external world.

Moore doesn't think that we can prove that the skeptic's argument is wrong; he just thinks that his argument is better. We might wish that we could do better than this against the external world skeptic. We might wish that instead of merely asserting a different argument, we could show what is *wrong* with the skeptic's argument. Moore doesn't try to do this (and doesn't think it can be done), but David Chalmers thinks we can show that there's something wrong the skeptic's argument. In particular, Chalmers will claim that the first premise of the skeptic's argument is false. Before turning to Chalmers's argument, however, we will take a look at a response to Moore's argument from one of the most famous philosophers of the 20th century—and, according to some, one of the most brilliant philosophers of all time, Ludwig Wittgenstein (1889-1951, pronounced VIT-gen-stein).

Wittgenstein's response to Moore

In *On Certainty* (which is really just a collection of his notes to himself), Wittgenstein reflects in on the concept of knowledge and how Moore uses (or misuses) it. Wittgenstein thinks that the fundamental flaw in Moore's response to the skeptic is the mistaken conception of knowledge on which his response depends. The basic problem with Moore's response, Wittgenstein thinks, is that he equates *subjective certainty* with *knowledge*. Just because a person is certain of something does not automatically mean that they know it. Sometimes we can be certain of things that turn out to be false (and thus we didn't really know them). There are plenty of **counterexamples** of people being very sure of things that aren't true (and thus they don't really know). For example, someone in the grips of a schizophrenic delusion might be very, very sure that they are Jesus Christ reincarnated—or the creatress of the universe who is temporarily a nid. There are lots of things about which an individual could feel very certain and yet not really know them. So the fact that Moore is very certain about the truth of the proposition "here is a hand" does not mean that he knows it. For Wittgenstein, one's level of subjective certainty about x doesn't establish that one knows x.

> To whom does anyone say that he knows something? To himself, or to someone else. If he says it to himself, how is it distinguished from the assertion that he is sure that things are like that? There is no subjective sureness that I know something. The certainty is subjective, but not the knowledge. So if I say "I know that I have two hands", and that is not supposed to express just my subjective certainty, I must be able to satisfy myself that I am right. But I can't do that, for my having two hands is not less certain before I have looked at them than afterwards. But I could say: "That I have two hands is an irreversible belief." That would express the fact that I am not ready to let anything count as a disproof of this proposition (*On Certainty*, 245).

So what is knowledge, if not subjective certainty? For Wittgenstein, to know something is to be able to rule out doubts concerning that thing. For example, if Ana said that Sara was at Dagwood's yesterday and I said, "How do you know that?" then whether or not Ana really knew that Sara was at Dagwood's yesterday would depend on her ability to rule out specific doubts that I raised. For example, Ana could say, "Because I was at Dagwood's yesterday too and I talked to her there." But, Wittgenstein asks, can Moore rule out doubts raised against the proposition "this is my hand"? In everyday discourse no one would raise such doubts, but in the context of philosophy the skeptic is raising exactly that kind of doubt—perhaps you are dreaming (or in a simulation). But can this doubt be ruled out? Wittgenstein thinks not and Moore actually agrees with him on this. Their dispute is whether or not my claim to know something depends on my ability to answer specific doubts regarding that thing.

Wittgenstein thinks that Moore's mistake is making knowledge into a *private, subjective* kind of thing, when in fact, according to Wittgenstein, knowledge is a very *public, objective* kind of thing. You cannot assure yourself that you know something just by believing it really hard (*On Certainty*, 245, 550). Rather, knowledge requires a kind of back and forth discourse between two different parties; it is what Wittgenstein famously called a language game. The **language game** involving the word "knowledge" involves being able to answer specific doubts raised about a claim. When those doubts have been answered, we call this "knowledge"; when they have not, then it is mere belief (even if it is a strongly held belief).

So how would Wittgenstein answer the external world skeptic? In contrast to Moore, Wittgenstein doesn't think that our most basic beliefs (like "this is my hand") can be justified and, thus, they do not count as knowledge. Rather, such beliefs are simply our starting points, our **axioms**. In mathematics, axioms cannot themselves be justified; rather, they are what the whole system depends on. Likewise, Wittgenstein thinks that our language games depend on certain propositions that cannot themselves be justified. This is the biggest contrast between Moore and Descartes, on the one hand, and Wittgenstein, on the other: Moore wants there to be a rational foundation for our knowledge about the world. He thinks that the most basic propositions of our language are also pieces of knowledge and that because of this the whole system of knowledge is a rational system. Moore's "here is a hand" argument is supposed to be the kind of thing that justifies the most basic propositions of our language—such as that there is an external world of physical objects. Wittgenstein, in contrast, does not think that the most basic assumptions of our language can be justified. Rather, they are just things we do or assume but that cannot themselves be given a rational justification. Here is a very telling passage from Wittgenstein's *On Certainty* where he makes this point:

> If I say "Of course I know that that's a towel" I am making an utterance. I have no thought of a verification. For me it is an immediate utterance. I don't think of past or future. (And of course it's the same for Moore, too.) It is just like directly taking hold of something, as I take hold of my towel without having doubts. And yet this direct taking-hold corresponds to a sureness, not to a knowing (*On Certainty*, 510-511).

For Wittgenstein, the most fundamental, basic propositions of our language— the ones that the external world skeptic is trying to cast doubt on—cannot properly be said to be "known." That is, they aren't pieces of knowledge.

Rather, they are just something we assume within a system—they are just something we *do*. Like taking hold of a towel that is before me, I don't even think about whether this is a towel or how I know it is, I just assume it is—or *act* as if it is. So Wittgenstein does not think that propositions like "there is an external world that exists outside of my perceptions" is something that can be known or proven. But the fact that we cannot justify the most basic propositions of our language isn't a defect in our language (or our knowledge); it's just the way things are and must be based on how language works. Just as it isn't a defect in a mathematical system (e.g., geometry) that the axioms themselves cannot be justified, so too it isn't a defect of our language that the axioms (such as that there is an external world) cannot be justified. Moore wants to be able to prove those axioms; Wittgenstein thinks that they can't be proven.

So is Wittgenstein an external world skeptic? Not really. Rather, he thinks that both the skeptic and Moore are making the same kind of mistake regarding what knowledge is. Both Moore and the skeptic accept the claim that *if we cannot justify our most basic, fundamental claims, then all of our knowledge about the external world is thrown into question*.[3] Wittgenstein, in contrast, rejects this proposition. To have knowledge is to be able to answer specific doubts raised about the claims one is making and one doesn't have to answer the external world skeptic in responding to specific doubts. For example, in my earlier example Ana didn't have to consider possible responses to Descartes dream argument in order to satisfy my doubts regarding Sara's whereabouts.

Likewise, I don't have to answer the external world skeptic in order to know that it was a brick that broke my window, for example. Moore and Descartes think that we need an answer to the skeptic in order to have knowledge about the world; Wittgenstein rejects this assumption. Wittgenstein's position is sometimes called **anti-foundationalism** because he rejects the **foundationalist** assumption (shared by Moore and Descartes) that for something to be known it must be justified by some other piece of knowledge. Wittgenstein thinks that the most basic assumptions of our discourse cannot themselves be justified and are not properly called knowledge. This might seem like a kind of skepticism, but if it is, it is very different from external world skepticism. The external world skeptic takes the fact that we cannot justify the most basic claims (such as that I am not dreaming or am not living in a computer simulation) as undermining all the rest of our knowledge of the external world. Wittgenstein agrees that the most basic claims cannot be justified but disagrees that this undermines the rest of our knowledge. As we will see, David Chalmers's view has some affinities with Wittgenstein's on this point.

---

3. Where Moore and the skeptic part ways is that Moore rejects the consequent of the conditional whereas the skeptic embraces the antecedent of the conditional.

Chalmers's response to external world skepticism

Invoking the Wachowskis' movie, *The Matrix*, the contemporary philosopher David Chalmers asserts that "even if we are in a matrix, our world is no less real than we thought it was. It just has a surprising fundamental nature."[4] This, in a nutshell, is Chalmers's response to external world skepticism. Let's unpack it to see what he's getting at.

Recall that one of the skeptical scenarios we've been contemplating (and that is an oft-used example in contemporary philosophy) is that of a computer simulation—for all you know, your whole life has been lived within a computer simulation. The skeptic's point is that if that were true, then there wouldn't be an external world of the sort you think (one of physical objects in time and space). The skeptic also thinks that you can't know that you aren't in a computer simulation. Chalmers does not try to dispute the claim that we can't know that we aren't in a simulation. Rather, he disputes the first claim—that *if we were*, then all our knowledge claims (that implicitly assume an external world) would be mistaken. Chalmers thinks that this doesn't follow. Here's why.

Consider the example of **Eddington's two tables**. Arthur Eddington was a famous physicist who in a 1927 lecture at University of Edinburgh gave the following example. Consider a kithen table as revealed to us *by science*: it is composed of mostly empty space (since an atom is mostly empty space), is in constant flux (since atoms are), has no colors (since atoms are not colored), etc. This is in stark contrast to the kitchen table as revealed *by our senses* which is solid (thus not empty space), stable (thus not in constant flux), and has a specific color (in my case, cream-colored). But does the existence of the "scientific table" undermine our "common sense" table? It would be absurd to think that it does.

Rather, what science reveals as true about the table (and all other objects, too) is just something surprising about the table's underlying nature. The table is still the common sense table; it is just revealed to have a surprising underlying nature. Thus, the table's surprising underlying nature doesn't make the table not a table anymore. It's still a table and we can still properly describe it as being solid, stable, and cream-colored.[5]

Now consider what would have to be the case if the reality in which we were currently living was in fact a computer simulation. That reality would still have a physics, a chemistry, a biology, etc.—in short, it would have all of the levels of reality that we think it would. But it would also have one further level of reality that we don't think it has: it would have a level of reality *lower* than physics—the computational level of the computer

---

4. David Chalmers, The Matrix as Metaphysics in *Philosophers Explore the Matrix*, Oxford University Press, 2005.

5. There is actually a deep philosophical (metaphysical) question about the relationship between the different levels of description which we won't get into here. But this metaphysical question doesn't seem to undermine the claim that we can properly describe the table as solid, etc.

program. Physics (think electrons, quarks, etc.) is typically thought of as the lowest level at which we can study reality.

Physics applies to everything that exists (well, assuming only physical things exist), whereas biology applies only to some things. That is, the laws of physics apply to everything whereas the laws of biology apply only to some things. For example, a frog is a biological and physical thing, whereas a rock is a physical thing and hence biological rules do not apply to it. If we were in a simulation right now, there would still be frogs and there would still be biological organisms and physical realities, it's just that those physical things would also have a further, surprising underlying nature: they would be also be, at root, computational entities—that is, they would really be 1s and 0s of a computer program. But does that mean they wouldn't be physical objects? No, it doesn't. Rather, the frog in a computer simulation would still be a biological and physical organism, it would just also be a computational object. This would be surprising (if we didn't think we were living in a simulation), but it wouldn't make the frog not a frog.

So Chalmers's response is specifically about matrix-like scenarios—i.e., scenarios in which we are living in a computer simulation. His argument contrasts with what the external world skeptic would have us believe, the fact that we were living in a computer simulation does not mean that there isn't an external world filled with objects that have roughly the kinds of properties that we perceive them to have. Thus, Chalmers takes himself to have shown that there is a certain class of skeptical scenarios that actually don't do what the skeptic thinks they do. Sure, there are some beliefs that would be false if we were, in fact, living in a simulation. For example, the belief that physics is the most basic level of reality would be false in a simulation (since in a simulation it would be the computational level that would be the most fundamental). But these beliefs are more abstract philosophical/scientific beliefs rather than every day beliefs. Since the skeptic takes the skeptical scenarios to undermine all of our knowledge about the external world, the skeptic is mistaken.

Chalmers's response relies on a particular account of how our words apply to things, what is called the causal theory of reference. The basic idea is that the way our words attach to (refer to) things in the world is by means of our causal histories with those objects; we don't have to understand the complex natures of those objects (e.g., that water is H2O) as long as those are the objects we have always used our words to refer to. So, for example, if I have always lived in a computer simulation and in that simulation I have always used the word "frog" to refer to a particular kind of object in that simulation (i.e., the frog simulations) then that is what the term "frog" means for me. On the other hand, if I am not living in a simulation then the term "frog" refers to objects whose fundamental nature is physical. But here is where a little wrinkle arises in Chalmers's neat argument: suppose that I have learned the term "frog" in a non-simulation world but then one night in my sleep I am kidnapped and put into a simulation that is identical (from the point of view of my experience) to the world as I've always known it. In that case, the skeptical problem would return in full force because now none of my words (which are anchored to the non-simulated reality in which I learned the terms) refer to anything in my new simulated reality and thus almost all of my beliefs and assertions are now false. So when I

now say "there is a frog" when the froggish thing hops by, that statement is false because the meaning of "frog" for me is tied to my former reality—the non-simulated reality that I falsely think I'm still in.[6] In light of this, Chalmers's victory against the external world skeptic is a limited one since it applies only to certain skeptical scenarios but not others. If I am and have always been in a simulation, then Chalmers's response to the skeptic is successful. But as we have just seen there are other kinds of skeptical scenarios where Chalmers's response is not successful and the external world skeptic wins. So Chalmers's response is not an unmitigated defeat of external world skepticism.

Study questions

1. True or false: Descartes does not think that he can be certain of anything.
2. True or false: Moore thinks that he can prove that he isn't dreaming (or in a skeptical scenario).
3. True or false: Moore's response to external world skepticism is direct rather than indirect.
4. True or false: Wittgenstein thinks that Moore is correct to say the he knows that "this is a hand."
5. True or false: Foundationalists claim that in order for some piece of knowledge to be justified, it must be justified by some other piece of knowledge
6. True or false: Wittgenstein and Moore agree that we can know basic propositions like "this is my hand."
7. True or false: Wittgenstein thinks that as long as one is subjectively certain about something, one knows it.
8. True or false: Chalmers thinks that if we were living in a simulation, then most of our beliefs about the external world would be false.
9. True or false: Eddington's two tables is supposed to show that physics proves that there aren't really tables.
10. True or false: Chalmers thinks that his defeat of external world skepticism applies equally well to any skeptical scenario.

---

6. Note that the same would be the case if I had always lived in the simulated reality and then were somehow moved to the non-simulated reality.

CHAPTER 2: THE PROBLEM OF OTHER MINDS

THE PROBLEM OF OTHER MINDS

Matthew Van Cleave

Have you ever wondered whether another person might see colors in a radically different way from you—or perhaps that you see colors in a radically different way from everyone else in the world? You might think that we could easily clear up any difference in perception by asking each other questions such as, "Is this green?" But this isn't what the original question is getting at. Rather, the original question is envisioning the possibility that your red is my green and vice versa. So we all call things by the same color—for example, we all refer to stop signs as "red" and grass as "green"—but what you in your head experience as red, I experience as green. It's just that I've learned to call things I experience as green "red" and things I experience as red, "green." Notice that if this were the case, it doesn't seem like there's any way that we could ever *know* that our subjective experiences are not the same. Philosophers refer to this possibility as the **inverted spectrum**. The inverted spectrum envisions a situation where our minds seem to differ radically without us being able to know that they do because we all describe objects in the world the same way but experience them differently. The inverted spectrum draws on a fundamental asymmetry about our own minds versus other minds: we seem to experience our conscious experience in a direct and unmediated way, whereas we experience the conscious experience of others only as mediated through their behaviors, including, importantly, what they *say*. The inverted spectrum further recognizes the possibility that there could be multiple different kinds of conscious experience that correspond to the same, exact behaviors. One way of putting this is that there is a many-to-one relationship between qualities of one's conscious experience—green, red, blue, yellow—and the behaviors that those experiences cause, such as saying "green."

Like the inverted spectrum, the problem of other minds draws on these same two fundamental claims about minds: 1) that we can only have direct, unmediated access to our own minds and 2) that there can be radically different causes of the very same intelligent behaviors. Developing this second claim will lead us to the problem of other minds. In order to do this we will take a little excursion into artificial intelligence and science fiction. Would it be possible for a machine to pass for a person? Could a machine behave in such a human-like way, that we mistake it as a human? The field of artificial intelligence has for many years now taken this question seriously and most within that field take the answer to be "yes." The idea is that we can engineer intelligence, from the ground up. If you think about it, human intelligence is manifest in certain kinds of behaviors. For example, being able to follow the instructions in setting up a camping tent or in putting together a piece of IKEA furniture. But perhaps the most prototypical intelligent behavior is the use of language. Could we teach a machine to respond intelligently with language—for example, to have a conversation? In 1950, the famous mathematician Alan Turing put forward a way of answering this question—the Turing Test. Turing proposed to reduce the question, "Can a machine think?" to the question of whether a machine could trick a human

being into thinking that it (the machine) was another human being, rather than a machine. Turing imagined a human participant communicating with either a computer or a human being via a keyboard; the participant's task was to determine whether or not she was communicating with a machine or a human. In order to do so, the participant could ask any question they liked, including questions like the following: "Imagine a capital letter "D" and rotate it 90 degrees to the left. Now place the "D" atop a capital letter "J." What object does this remind you of?"[1] If the machine were able to make the human think that it was actually a human (and not a machine), then, Turing claims, we should consider that machine to be intelligent. Turing thought that machines would eventually be able to pass this test—they could trick a human into thinking it was not a machine—and that therefore machines could think.

But there another interpretation of what's going on here, which is that such a machine would only give the *illusion* that it was thinking. That is, such a machine would *behave as if* it were thinking, but it wouldn't really be thinking because it was really nothing other than an ingenuously designed mechanism that behaved in a way that was indistinguishable from things (like humans) that can truly think. The idea is that humans have minds (and thus thoughts) but that machines don't have minds (and thus don't really think). Rather, machines only behave as if they are thinking; they appear intelligent, but really aren't.

But now notice where this response puts us. We have admitted that there could be things that behave indistinguishably from intelligent beings but that lack minds. If this is so, then it raises the question: how do we know that human beings other than ourselves have minds? For all we know, other human beings could be nothing other than being who behave intelligent (they act—speak and behave—as we would expect intelligent beings to act) but they don't really have minds. Rather, what drives their intelligent behavior is nothing other than an ingenuous mechanism. Of course, in one's own case, one knows that one has thought because one can observes one's own thoughts directly, it seems. But in the case of others, we can only infer the existence of their minds/thoughts through their behaviors. And since we have noted that sometimes the very same intelligent behaviors could be caused by things without minds (for example, computer programs), we cannot rule out that in the case of other people the same sort of thing is occurring—their speech and behavior is caused by ingenuously designed, but ultimately mindless, mechanisms. Perhaps you are unique in the universe in that only you have thoughts; everyone else is just a mindless machine, like Turing's computer program that passes the Turing Test. This skeptical scenario—that you are the only thing in the universe that thinks and has a mind—is what philosophers call **solipsism**. Solipsism is the skeptical scenario that defines **the problem of other minds**. The problem of other minds is simply that you cannot rule out solipsism; all of your experience is consistent with the possibility that you are the only thing in the universe with a mind. Of course, no one actually believes this but the skeptic's point is that you cannot rule out this seemingly absurd possibility. Here is the skeptic's **reconstructed argument**:

---

1. The correct answer is: an umbrella. This example comes from Daniel Dennett.

1. If we cannot rationally rule out that solipsism is true, then we cannot know that other people have minds.
2. We cannot rationally rule out that solipsism is true.
3. Therefore, we cannot know that other people have minds.

This argument is a **valid argument**—that is, the conclusion of the argument *must* be true if the premises are true. Thus, if there's a flaw in the argument it would have to be that one of the premises is false. As we will see in the following section, philosophers have taken aim at both premises and attacked the argument in very different ways.

There is another way of presenting the problem of other minds that will help us to zero in on an important aspect of what "mind" means in the context of the problem of other minds. This way of presenting the problem relies on the concept of a **philosophical zombie**. A philosophical zombie is an imaginary creature that behaves indistinguishably from a normal human being but who lacks any conscious experience of the world. A philosophical zombie is, so to speak, dead inside—it doesn't experience anything. Of course, a philosophical zombie won't know it is a philosophical zombie. It will answer questions like the above "D"-umbrella question correctly and it will be able to describe the fragrance of a rose and distinguish that smell from the smell of coffee. It will be able to talk about food it likes and the things that turn it on sexually. That is, it will behave and speak in a way that is indistinguishable, from an outside perspective, from a normal human being who does has these different conscious experiences. But its insides will be like the insides of a rock—there's nothing there. This isn't to say, of course, that there isn't a complex mechanism (the brain) that is causing all of these intelligent behaviors. Rather, it is just to say that there is no conscious experience attached to the mechanics of the brain. The neurons do what they do in causing the behaviors; there just isn't any conscious experience connected with the functioning of those mechanisms.[2] Thus, we can restate the problem of other minds in terms of the skeptical scenario in which everyone in the world except you is a philosophical zombie. How could you know this isn't the case? Again, it isn't that the skeptic thinks that this is that case; rather, it's that the skeptic think you can rule it out and therefore that you can't know that other people have minds.

This way of setting out the problem makes is clear that the aspect of mind that is operative in the "problem of

---

2. Are philosophical zombies logically possible? One way of seeing that they are is by considering blindsight patients. Blindsight is a neurological disorder in which patients can correctly respond to questions about something in their visual field although they have no conscious experience of that field of vision. The reality (made clear in blindsight patients) that information can be conveyed from the visual system to the executing processes that govern behavior and speech without there being any conscious awareness of the properties conveyed through vision is an existence proof of the possibility that a philosophical zombie represents in the extreme.

other minds" is that of **conscious experience**. If all we meant by "mind" was simply whatever it is that causes a certain class of behaviors (namely, those we deem "intelligent"), then there is no problem of other minds. In that definition of "mind," the computer that passes the Turing Test has a mind and so does the philosophical zombie, since in both cases there is something that is causing those intelligent behaviors. Defining the mind in this way enables us to verify the existence of minds in a third-person type of way—that is, I can know that you have a mind because I can know that you engage in intelligent behaviors (such as conversing with me). However, if we focus on the conscious experience aspect of minds, then this seems to be something that can be known only in the first-person. That is, I alone have direct access to my own conscious experience; the conscious experiences of others can only be inferred through observations of others' intelligent behaviors. But since these intelligent behaviors could be caused by things that aren't conscious, I cannot confidently infer that others have conscious experience. That is the problem of other minds.

Ludwig Wittgenstein had his own way of addressing the problem of other minds. See if you can understand what Wittgenstein is saying in the following passage.

> "If I say of myself that it is only from my own case that I know what the word 'pain' means—must I not say that of other people too? And how can I generalize the one case so irresponsibly?
>
> Well, everyone tells me that he knows what pain is only from his own case!—Suppose that everyone had a box with something in it which we call a 'beetle.' No one can ever look into anyone else's box, and everyone says he knows what a beetle is only by looking at his beetle.
>
> Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing.— But what if these people's word 'beetle' had a use nonetheless?—If so, it would not be as the name of a thing. The thing in the box doesn't belong to the language-game at all; not even as a Something: for the box might even be empty" (*Philosophical Investigations*, §293).

Wittgenstein uses a metaphor of a "beetle in a box." What are the beetle and the box metaphors for? (If you don't know, reread it and think about it before proceeding.) *The box* is a metaphor for the mind—this is why we cannot see into others' boxes; other people's minds are "black boxes." This is Wittgenstein's colorful way of putting forward the Cartesian[3] view of mind on which the problem of other minds depends. **The Cartesian view of mind**, as I use that term here, just refers to the idea that I can know my own mind in a direct and unmediated way, whereas I can only know the minds of others indirectly, through their behavior. *The beetle*

---

3. The term "Cartesian" is just a way of referring to the ideas of Rene Descartes, the 17th century philosopher who was introduced in the chapter on external world skepticism. So the "Cartesian view of mind" is a just a way of referring to the concept of the mind that Descartes held.

is simply a way of referring to an individual's conscious thoughts and sensations. So I can see my own beetle/thoughts because it exists in my own box/mind, but I cannot see others' beetles/thoughts because they exist in others' boxes/minds. Now look at what Wittgenstein says in the last (third) paragraph of the above passage: it is "quite possible for everyone to have something different in his box." Do you see how that is just a colorful, metaphorical way of raising the inverted spectrum problem? Now look at the very last line: Wittgenstein envisions that the box could be empty and we would be none the wiser. Do you see how this is a metaphorical way of raising the problem of other minds? People might be talking intelligently about their beetles/thoughts even if there really aren't any beetles/thoughts. Wittgenstein's point is that we (our language) is not able to penetrate into the minds of others. That we cannot do so is what gives rise to the problem of other minds. We cannot know that the words people speak have any thoughts/sensations tied to them. We can know in our own case but we cannot know in the case of others. Wittgenstein actually has his own kind of solution to the problem of other minds that he hints at here, but we will save a discussion of that until the next section when we consider some of the famous attempts to solve the problem of other minds.

Study questions

1. True or false: If someone's color spectrum were inverted relative to mine (and had always been so since they were an infant), I would be able to discover this by asking them questions such as "is this thing red or green?"
2. True or false: The Turing Test enables us to discover whether or not someone has conscious thoughts and sensations.
3. True or false: The Turing Test was conceived as a test that would enable us to answer the question: "Can a machine think?"
4. True or false: The aspect of "mind" that is operative in the problem of other minds is that of conscious experience.
5. True or false: Wittgenstein's "beetle in the box" can be seen as a metaphorical way of describing *both* the inverted spectrum and the problem of other minds.

For deeper thought

1. If we were to define the mind as "the thing that causes intelligent behaviors" then would there still be a problem of other minds? Why or why not?
2. Does the skeptic think that other people *don't* have minds? Why or why not?
3. What is the "Cartesian view of mind"?
4. Even if I can't know others' minds directly, can't I still know them by inferring based on their behaviors? For example, if someone is crying and I ask why and they say that their grandfather just passed away, can't I rationally conclusion that they are sad? And if it is rational to say that they are sad and if sadness

is a thought or sensation that require a mind, then isn't is rational to say that they have a mind? What might the skeptic say in response to this reasoning?

5. Suppose a neuroscientist were to say: "I can solve the problem of other minds, just let me show you inside the brain and you can see all these billions of neurons doing their thing. Since the mind is the same thing as the brain, I have just proved to you the existence of other minds (since this brain is something you can observe in action and it is not *your* brain)." How might the skeptic respond to the neuroscientist?

Responses to other minds skepticism

We will consider two types of responses to other minds skepticism: the "analogical inference" solution and the behaviorist solution. The analogical inference solution that we will consider comes from the 20th century philosopher, Bertrand Russell (1872-1970). Here is Russell's reconstructed argument:

1. I know through introspection that in my own case, conscious mental states (M) regularly precede many of my intelligent behaviors (B).
2. I can observe similar kinds of intelligent behaviors (B), in others.
3. Therefore, based on the similarity of (B) in me and (B) in others, (B) in others are probably regularly preceded by conscious mental states (M).[4]

Suppose that I am working on a tough math problem, or figuring out how to respond carefully to a sensitive issue in a text message, or figuring out what my next move should be in a game of chess, or saying "that smells wonderful" after smelling a lilac bush. In all of these cases I am often aware of the conscious thoughts I am having and I recognize that my intelligent behavioral responses are regularly preceded by those conscious thoughts. Notice that whereas my conscious thoughts are something that only I have access to, the behaviors are something that I can equally well observe in both myself and others.[5] Since in my own case I can observe

---

4. I should note that Russell himself argues something stronger than what I've presented here. Russell thinks that we can discover in our own case that M causes B. But I have presented the weaker claim that M and B are correlated since the weaker claim is all that is needed to support the inference to other minds. Russell's claim that M causes B (and that this is something he can know via introspection) is problematic. Philosophers will recognize that my reconstruction raises the problem of epiphenomenalism which Russell's formulation doesn't.

5. One way of putting this is that conscious experience is first-person accessible, whereas behaviors are third-person accessible. The astute reader should recognize that this is the Cartesian view of mind.

both my conscious thoughts, on the one hand, and the intelligent behaviors, on the other, I am able to establish a correlation between them: M → B (read this as "M reliably precedes B"). So, for many of the intelligent behaviors I observe (namely, my own), I can observe that they are preceded by conscious mental states. But I can also observe many intelligent behaviors of others. For example, if my friend Grace says, "that lilac bush smells wonderful" after smelling a lilac bush, then this is one of her intelligent behaviors that I can observe. Although I cannot observe her mental states (such as the smells that she smells), I know that when I make similar statements in similar situations, my behaviors correlate with mental states (for example, the delightful fragrance of a lilac bush) which I *do* have access to. Or, when Grace puts my king in check with her queen and then says "check," I cannot observe the conscious thoughts that she has when she does this, but I do know that when I do similar things (that is, when I put others into "check" in a game of chess), those behaviors are correlated with a series of conscious thoughts.

Russell thinks it is perfectly rational to assume, based on the correlation, M → B in my own case, that the same thing holds for others, based on the strength of analogy between our intelligent behaviors, which are very similar.

ME	Thought ⟶	Behavior
OTHERS	??? ⟶	Behavior
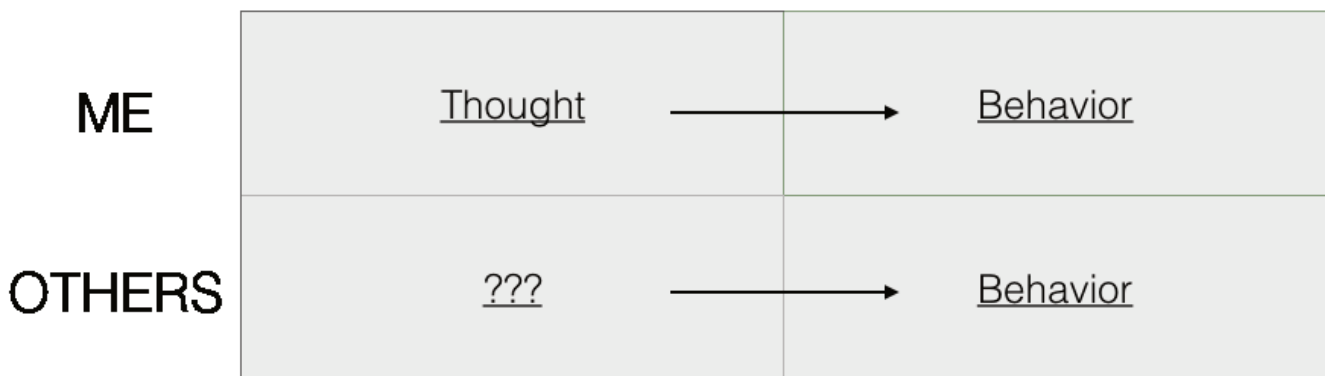
This figure depicts the analogy in Russell's **analogical argument**. The strength of the argument rests on the strength of analogy between my behaviors and the behaviors of others. Are they really very similar or are they quite different?

Russell's claim is that they are very similar and so we should think that they are similar in other respects, too. The problem of other minds says that we do not know whether others have conscious thoughts; Russell thinks we do know this because 1) our own intelligent behaviors are similar to the intelligent behaviors of others and that therefore, 2) we can generalize from our own case to the case of others.

However, a problem immediately arises for Russell's solution: if I have observed many more intelligent behaviors in others than in myself, isn't my generalization from my own experience a **hasty generalization**? A hasty generalization is an informal logical fallacy according to which one makes a generalization to a whole population based on too few examples. For example, if the first Michigander I met owned a canoe and based on that *one* Michigander I asserted, "All Michiganders own canoes," then that would be a hasty generalization.

I am generalizing to all Michiganders from my experience with only one Michigander. That does not seem like a very strong logical inference. Perhaps the Michigander I met was idiosyncratic and not representative of all Michiganders. The problem is that Russell seems to be committing exactly this fallacy in his solution to the problem of other minds. Suppose I have 100 friends, all of whom I believe have minds like mine and all of whose intelligent behaviors I have observed. I also have observed, in my *one instance*, the correlation, M →B. But how can I legitimately infer based on this one case that this correlation holds in all of the other cases (for example, that it holds in the case of all of my 100 friends)? That is a hasty generalization, *par excellence*. It is making the same mistake as inferring that all Michiganders own canoes from my experience with one Michigander who owned a canoe.

Here is the relevant passage in Russell where I think he makes the fatal error:

> "If, whenever we can observe whether A and B are present or absent, we find that every case of B has an A as a causal antecedent [in our own case], then it is probable that most Bs have As as causal antecedents [in the case of others], even in cases where observation does not enable us to know whether A is present or not."

I have added in bold text and square brackets the explicit assumption that Russell needs in order for his analogical argument to work. But when that assumption is made explicit, it is clear what the problem is: what entitles Russell to assume that my own case is the same as the case of others? To assume that they are would seem to commit the fallacy of **begging the question** against the other minds skeptic. And if we actually think of how the generalization works, it is a bad generalization—it is a hasty generalization. This is what Wittgenstein is getting at in the above passage when he says that in assuming that the meaning of "pain" is given by my subjective conscious experience, I "generalize the one case so irresponsibly." Wittgenstein is making the same point that I have been making (in addition to others): that generalizing from my own conscious experiences to the conscious experiences of others is a hasty generalization.

After all, for all I know everyone else is a philosophical zombie and I alone am unique in having conscious experiences. Without having further knowledge of other Michiganders, it would be totally irrational for me to infer that all Michiganders own canoes based on only the one case; without having further knowledge of others' minds, it would be totally irrational for me to infer that others have conscious mental states, based only on my own case. The problem is that whereas I can go and observe other Michiganders and their canoe- owning propensities, I cannot observe others' minds to determine whether they have conscious experience. Therein lies the problem of other minds.

I do not see how Russell can answer the above objection. It seems that either he committing a hasty generalization (if he is generalizing from one's own intelligent behaviors to the intelligent behaviors of everyone else) or he is begging the question (if he is assuming that all intelligent behaviors—my own and others—are similar in all respects). It is important to see that Russell accepts, and does not question, the Cartesian view of mind that gives rise to other minds skepticism. One might think that that is the problem: once one accepts

the Cartesian view of mind, the problem of other minds is unavoidable. This is what the behaviorist solution attempts to challenge. However, in an important sense, the behaviorist solution to the problem, isn't really a solution at all, but rather a rejection of the problem in the first place. The behaviorist thinks that the problem of other minds is only a problem because it assumes a mistaken view of the nature of the mind. Thus, I prefer to refer to the behaviorist solution as a dissolution of the problem, since it rejects the terms in which the problem is presented. Dissolutions of philosophical problems reject that the problem really is a problem. Thus, they attempt to show what the mistaken assumptions are that give rise to the problem. The behaviorist dissolution to other minds skepticism does this by rejecting the Cartesian view of mind on which other minds skepticism depends. In contrast, Russell's attempted analogical inference solution to other minds skepticism accepts the Cartesian view of mind on which other minds skepticism depends.

The behaviorist dissolution of other minds skepticism

According to the behaviorist tradition[6], the mind is as the mind does. The mind is not best conceived as a private domain that only the subject has access to, as the Cartesian view of mind holds. Rather, "mind" is just a fancy way of referring to a certain class of intelligent behaviors or to the input/output functions of the brain. The 20th century Oxford philosopher, Gilbert Ryle famously referred to the Cartesian view of mind as "the dogma of the Ghost in the Machine." This phrase captures the Cartesian idea that there was some private domain inaccessible to anyone except the subject herself ("ghosts") and that doesn't function according to the normal physical laws that govern physical things ("machines"). Minds, according to the Cartesian view, are mysterious, ghostly things that cannot be explained in terms of scientific principles. The behaviorist tradition seeks to overturn this view of the mind as something ultimately mysterious and seek to replace it with a view according to which the mind can be studied scientifically like any other object in the universe. If Cartesians are ghost hunters, behaviorists are Penn and Teller calling "bullshit."

As Ryle notes, the problem of other minds is a direct consequence of the Cartesian view of mind:

> "Not unnaturally, therefore, an adherent of the official theory finds it difficult to resist this consequence of his premises, that he has no good reason to believe that there do exist other minds than his own" (*Concept of Mind*, p. 3).
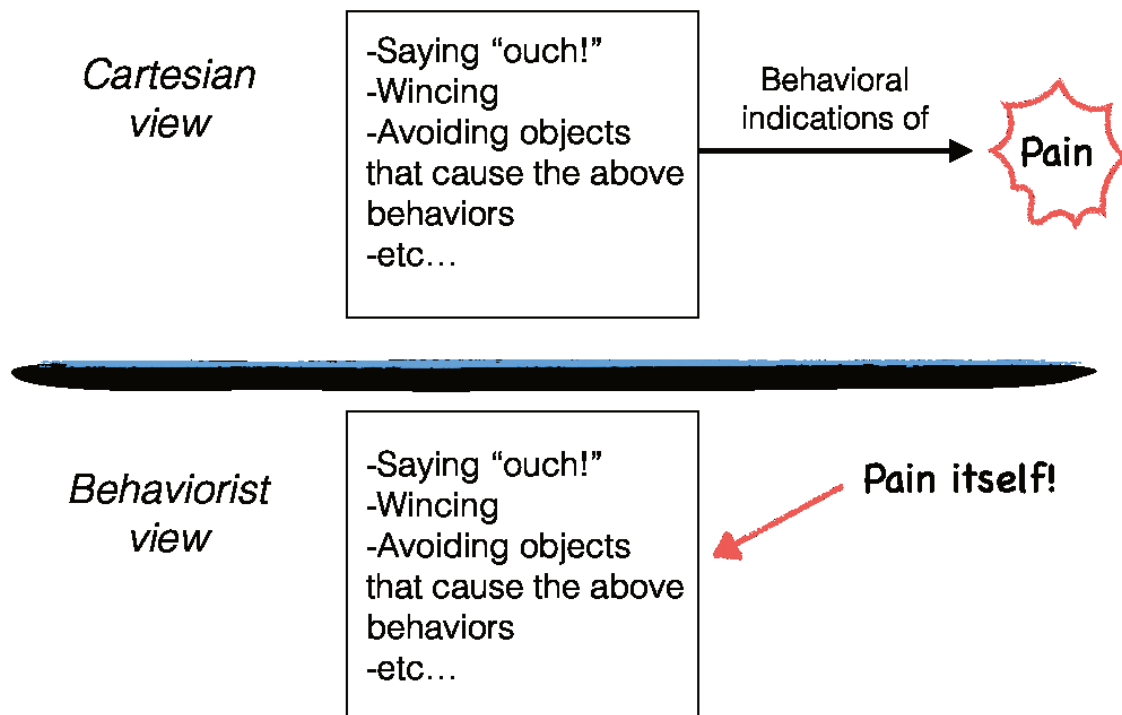
But there's really no problem if the mind isn't some private domain but, rather, something publicly accessible. The typical behaviorist way to try to convince people of this is to focus on language—to think about the

---

6. Here I include within the behaviorist tradition the heir of behaviorism, functionalism. Although there are important differences, both behaviorists and functionalists see the mind as essentially tied to actions. Whereas traditional behaviorists treat the brain as a kind of black box, functionalists, influenced by modern-day cognitive science, try to understand the input-output functions that the brain instantiates.

language we use to describe minds rather than to think about minds themselves. Behaviorists like Ryle and Wittgenstein were keen to analyze our language and to understand how language works because they believed that language was often something that misled us regarding the nature of things. Once we understand how language is meaningful, including what our words refer to, many traditional philosophical problems will be revealed as pseudo-problems—problems based on a misunderstanding rather than actual, deep problems. For behaviorists, the mind is nothing other than the way we refer to the mind using certain kinds of mental terms, such as pain, hope, fear, intelligence, and so on. The key to understanding the nature of the mind is to understand how these terms become meaningful to us—what they must refer to if we are to learn their meaning.

A good example of a behaviorist doing this is Wittgenstein in the "beetle in the box" passage cited earlier. It is clear in that passage that what Wittgenstein is attempting to understand is how we learn the meaning of a mental term like "pain." The traditional idea is that we learn the meanings of terms like "pain" by singling out a kind of experience in our mind's eye and then having that experience be the thing that our term "pain" refers to. Wittgenstein's point in that passage is that this can't be right because in that case it is possible that our words could refer to radically different things and yet we would never know it. It might even be the case that a person had no experience at all—no beetle in their box. Wittgenstein's point is that our terms don't get their meaning by referring to some inner, private domain of our conscious experiences. Rather, our mental terms get their meaning by how we use them and they always, in the end, refer to publicly observable phenomena.

Gilbert Ryle followed Wittgenstein's lead in this and put forward his own behaviorist conception of how our mental terms get their meaning. The figure below gives a nice comparison of the Cartesian and behaviorist accounts of the meaning of our mental terms—terms like "pain."

Cartesian view

-Saying "ouch!"
-Wincing
-Avoiding objects that cause the above behaviors
-etc…

Behavioral indications of → Pain

Behaviorist view

-Saying "ouch!"
-Wincing
-Avoiding objects that cause the above behaviors
-etc…

← Pain itself!

As you can see, for Ryle pain isn't some inner, hidden phenomena; rather, "pain" refers to certain kinds of observable behaviors—saying "ouch," wincing in pain, having the tendency to avoid objects that have caused pain in the past, and so on. These behavioral manifestations of pain just are what pain is for Ryle, whereas for the Cartesian they are merely behavioral indications of a private conscious experience. For the behaviorist, however, there's nothing further that the term "pain" could refer to—certainly not to any internal conscious experiences. After all, how could we ever learn the meanings of mental terms if they referred to things that our mothers couldn't show us? Wittgenstein's point was that there can't be because anything that is essentially private is not something that could figure into the meaning of a publicly shared language.

Ryle has a nice analogy that he used to illustrate the kind of thing that a mind is. Imagine a person who doesn't yet understand what the term "average taxpayer" means. This person is puzzled by the fact that they have never encountered this "average taxpayer" in real life. How mysterious a person this "average taxpayer" must be! They appear in all kinds of official documents talking about facts about the country, but they have never been encountered in real life! The mistake this person is making, of course, is that they are thinking that "average taxpayer" refers to some concrete particular object when in fact it refers to a more abstract object—to the economic average of a whole country. Particular taxpayers are things that we encounter in time and space (Bob, Sue, Sally) but "average taxpayer" is not that kind of thing. Ryle thinks that the Cartesian view of mind makes a similar mistake—what he calls a "category mistake"—in trying to locate the mind in the realm of **concrete particular objects**. The mind is not a concrete particular object, Ryle thinks, but a kind of **abstract object**, like "average taxpayer." "Mind" is simply a shorthand way of referring to a range of different mental terms that themselves designate various types of intelligent behaviors.

That doesn't mean that minds aren't real; it just means that we go awry, and thus lead ourselves into confusions, when we go looking for minds in the world of concrete particular objects.

How does the behaviorist conception of mind solve the problem of other minds? Fairly straightforwardly. Other minds skepticism gets its teeth from the idea that minds seem to be something that we have only first-person access to. This means that we can only know our own minds directly and can only know others' minds indirectly, based on their behaviors. Since those behaviors could be associated with very different inner experiences—including, crucially, *no experiences at all*—and we cannot rule out that this isn't the case, it follow that we cannot know others have minds. The behaviorist conception of mind rejects that minds are something that can only be known in the first-person. Rather, "minds" are a kind of abstract object that consists of an assemblage of all the different mental terms that we use. And since these mental terms refer to publicly observable phenomena, it turns out that we can know other minds just as directly as we can know our own minds!

What should we say about the behaviorist dissolution of the problem of other minds? Is it successful? It does seem that there's something that the behaviorist conception of mind neglects—the fact that we do, in our own case, have conscious experiences and that these conscious experiences seem to be an important part of what minds are. A philosophical zombie seems to be a radically different being than we are—even if we have no way of knowing whether someone is a philosophical zombie or not. One way of revealing what seems to be wrong with the behaviorist conception of mind is by way of a famous intellectual joke:

> Behaviorist #1 to Behaviorist #2: (after a romantic romp in bed): That was good for you, how was it for me?

The joke, of course, is that it seems absurd to not know how an experience like sex was for oneself. You don't need to observe one's own behaviors to answer this question. But it seems that this is exactly what the behaviorist is saying. We can make the same point about pain: if you hit your thumb with a hammer you don't have to observe yourself saying "ouch" and wincing in pain in order to know that you are in pain! Thus, the problem for the behaviorist view of mind is that it seems to neglect the reality of conscious experience. Even if people like Wittgenstein and Ryle are correct about the meaning of mental terms, it seems that we can nevertheless meaningfully ask the question about whether other people have conscious experiences like my own—or even have them at all! That this seems like a perfectly meaningful question one can ask about the world mitigates against the behaviorist's attempted dissolution of the problem of other minds.

Thus, if Russell's attempt to solve the problem of other minds skepticism fails (as I have argued) and if the behaviorist's attempted dissolution of other minds skepticism is also unsuccessful, then we are still without a solution to the problem of other minds. This doesn't mean that we haven't made progress, however, since sometimes knowing what doesn't work is part of how you get to a solution that does work. Whether or not there is a solution to the problem of other minds continues to be something that philosophers debate.

And the one of the most fundamental divides within this debate turns on the two views of mind that I have introduced above: the Cartesian and behaviorist views of mind. These two camps mark radically different ways of approaching a range of different philosophical questions the concern the mind.

Study questions

1. True or false: Russell's analogical inference solution to other minds skepticism accepts the Cartesian view of mind.
2. True or false: Russell thinks that since we can know our own minds directly and can correlate our conscious experiences with our intelligent behaviors, we can rationally infer that others also have conscious experiences attached to their intelligent behaviors (even though we can't observe those conscious experiences).
3. True or false: The Cartesian view of mind and the behaviorist view of mind agree that we mental terms like "pain" refer to the publicly observable behaviors, such as someone yelling "ouch!"
4. True or false: Behaviorists attempt to reorient philosophical questions about the mind (such as the problem of other minds) from the *mind itself* to how we *talk about* the mind.
5. True or false: According to Ryle, minds are kind of concrete particular object, albeit one that isn't physical—like a ghost.

For deeper thought

1. What is the difference between concrete particular objects and abstract objects? Give an example of each type of object.
2. What is the difference between a solution to a philosophical problem and a dissolution of a philosophical problem?
3. Can you think of a better solution to the problem of other minds—one that doesn't encounter any of the problems of the above solutions? That is, is there a way of acknowledging the reality of conscious experience but that doesn't lead to other minds skepticism?
4. Why does Wittgenstein think that inner conscious experiences are irrelevant to the meaning of mental terms like pain? Do you think he is correct about this? Why or why not?
5. According to Ryle, how is the term "average taxpayer" similar to the term "mind"?

CHAPTER 3: THE MIND-BODY PROBLEM

THE MIND-BODY PROBLEM

Matthew Van Cleave

Introduction: A pathway through this chapter

What is the relationship between the mind and the body? In contemporary philosophy of mind, there are a myriad of different, nuanced accounts of this relationship. Nonetheless, these accounts can be seen as falling into two broad categories: dualism and physicalism.[1] According to **dualism**, the mind cannot be reduced to a merely physical thing, such as the brain. The mind is a wholly different kind of thing than physical objects. One simple way a dualist might try to make this point is the following: although we can observe your brain (via all kinds of methods of modern neuroscience), we cannot observe your mind. Your mind seems inaccessible to third-person observation (that is, to people other than you) in a way that your brain isn't. Although neuroscientists could observe activation patterns in your brain via functional magnetic resonance imagining, they could not observe your thoughts. Your thoughts seem to be accessible only in the first person—only you can know what you are thinking or feeling directly. Insofar as other can know this, they can only know it indirectly, though your behaviors (including what you *say* and how you *act*). Readers of previous chapters will recognize that dualism is the view held by the 17th century philosopher, René Descartes, and that I have referred to in earlier chapters as the **Cartesian view of mind**. In contrast with dualism, **physicalism** is the view that the mind is *not* a separate, wholly different kind of thing from the rest of the physical world. The mind is constituted by physical things. For many physicalists, the mind just is the brain. We may not yet understand how mind/brain works, but the spirit of physicalism is often motivated by something like **Ockham's razor**: the principle that all other things being equal, the *simplest* explanation is the *best* explanation. Physicalists think that all mind related phenomena can be explained in terms of the functioning of the brain. So a theory that posits both the brain and another *sui generis* entity (a nonphysical mind or mental

---

1. Readers who are familiar with the metaphysics of minds will notice that I have left out an important option: **monism**, the idea that there is ultimately only one kind of thing in the world and thus the mental and the physical do not fundamentally differ. Physicalism is one version of monism, but there are many others. Bishop George Berkeley's idealism is a kind of monism as is the panpsychism of Leibniz and Spinoza. I have chosen to focus on physicalism for pedagogical reasons, because of its prominence in contemporary philosophy of mind, because of its intuitive plausibility to those living in an age of neuroscience, and because the nuances of the arguments for monism are beyond the scope of this introductory treatment of the problem.

properties) violates Ockham's razor: it posits two kinds of entities (brains *and* minds) whereas all that is needed to explain the relevant phenomena is one (brains).

The mind-body problem is best thought of not as a single problem but as a set of problems that attach to different views of the mind. For physicalists, the mind-body problem is the problem of explaining how conscious experience can be nothing other than a brain activity—what has been called "**the hard problem**." For dualists, the mind-body problem manifests itself as "**the interaction problem**"—the problem of explaining how nonphysical mental phenomena relate to or interact with physical phenomena, such as brain processes. Thus, **the mind-body problem** is that no matter which view of the mind you take, there are deep philosophical problems. The mind, no matter how we conceptualize it, seems to be shrouded in mystery. That is the mind-body problem. Below we will explore different strands of the mind-body problem, with an emphasis on physicalist attempts to explain the mind. In an era of neuroscience, it seems increasingly plausible that the mind is in some sense identical to the brain. But there are two putative properties of minds—especially human minds—that appear to be recalcitrant to physicalist explanations. The two properties of minds that we will focus on in this chapter are "original intentionality" (the mind's ability to have meaningful thoughts) and "qualia" (the qualitative aspects of our conscious experiences).

We noted above the potential use of Ockham's razor as an argument in favor of physicalism. However, this simplicity argument works only if physicalism can explain all of the relevant properties of the mind. A common tactic of the dualist is to argue that physicalism *cannot* explain all of the important aspects of the mind. We can view several of the famous arguments we will explore in this chapter—the "Chinese room" argument, Nagel's "what is it like to be a bat" argument, and Jackson's "knowledge argument"—as manifestations of this tactic. If the physicalist cannot explain aspects of the mind like "original intentionality" and "qualia" then the simplicity argument fails. In contrast, a tactic of physicalists is to either try to meet this explanatory challenge or to deny that these properties ultimately exist. This latter tactic can be clearly seen in Daniel Dennett's responses to these challenges to physicalism since he denies that original intentionality and qualia ultimately exist. This kind of eliminativist strategy, if successful, would keep in place Ockham simplicity argument.

Representation and the mind

One aspect of mind that needs explaining is how the mind is able to represent things. Consider the fact that I can think about all kinds of different things— about this textbook I am trying to write, about how I would like some Indian food for lunch, about my dog Charlie, about how I wish I were running in the mountains right now. Medieval philosophers referred to the mind as having **intentionality**—the curious property of "aboutness"—that is, the property of an object to be able to be *about* some other object. In a certain sense, the mind seems to function kind of like a mirror does—it reflects things other than itself. But unlike a mirror, whose reflected images are not inherently meaningful, minds seem to have what contemporary philosopher

John Searle calls "**original intentionality**." In contrast, the mirror has only "**derived intentionality**"—its image is meaningful only because something else gives it meaning or sees it as meaningful. Another thing that has derived intentionality is words, for example the word "tree." "Tree" refers to trees, of course, but it is not as if the physical marks on a page inherently refer to trees. Rather, human beings who speak English use the word "tree" to refer to trees. Spanish speakers use the word "arbol" to refer to trees. But in neither case do those physical marks on the page (or sound waves in the air, in the case of spoken words) inherently mean anything. Rather, those physical phenomena are only meaningful because a human mind is representing those physical phenomena as meaningful. Thus, words are only meaningful because a human mind represents them in a meaningful way. Although we speak of the word itself as carrying meaning, this meaning has only derived intentionality. In contrast, the human mind has original intentionality because only the mind is the ultimate creator of meaningful representations. *We can explain the meaningfulness of words in terms of thoughts, but then how do we explain the meaningfulness of the thoughts themselves?* This is what philosophers are trying to explain when they investigate the representational aspect of mind.

There are many different attempts to explain what mental representation is but we will only cursorily consider some fairly rudimentary ideas as a way of building up to a famous thought experiment that challenges a whole range of physicalist accounts of mental representation. Let's start with a fairly simple, straightforward idea—that of mental images. Perhaps what my mind does when it represents my dog Charlie is that it creates a mental image of Charlie. This account seems to fit our first person experience, at least in certain cases, since many people would describe their thoughts in terms of images in their mind. But whatever a mental image is, it cannot be like a physical image because physical images require interpretation in terms of something else. When I'm representing my dog Charlie it can't be that my thoughts about Charlie just are some kind of image or picture of Charlie in my head because that picture would require a mind to interpret it! But if the image is suppose to represent the thing that has "original intentionality," then if our explanation requires some other thing that has that has original intentionality in order to interpret it, then the mental image isn't really the thing that has original intentionality. Rather, the thing interpreting the image would have original intentionality. There's a potential problem that looms here and threatens to drive the mental image view of mental representation into incoherence: the object in the world is represented by a mental image but that mental image itself requires interpretation in terms of something else. It would be problematic for the mental image proponent to then say that there is some other inner "understander" that interprets the mental image. For how does this inner understander understand? By virtue of another mental image in its "head"? Such a view would create what philosophers call an **infinite regress**: a series of explanations that require further explanations, thus, ultimately explaining nothing. The philosopher Daniel Dennett sees explanations of this sort as committing what he calls "**the homuncular fallacy**," after the Latin term, *homunculus*, which means "little man." The problem is that if we explain the nature of the mind by, in essence, positing another inner mind, then we haven't really explained anything. For that inner mind itself needs to be explained. It should be obvious why positing a further inner mind inside the first inner mind enters us into an infinite regress

and why this is fatal to any successful explanation of the phenomenon in question—mental representation or intentionality.

Within the cognitive sciences, one popular way of understanding the nature of human thought is to see the mind as something like a computer. A computer is a device that takes certain inputs (representations) and transforms those inputs in accordance with certain rules (the program) and then produces a certain output (behavior). The idea is that the computer metaphor gives us a satisfying way of explaining what human thought and reasoning is and does so in a way that is compatible with physicalism. The idea, popular in philosophy and cognitive science since the 1970s, is that there is a kind of **language of thought** which brain states instantiate and which is similar to a natural language in that it possesses both a grammar and a semantics, except that the representations in the language of thought have original intentionality, whereas the representations in natural languages (like English and Spanish) have only derived intentionality. One central question in the philosophy of mind concerns how these "words" in the language of thought get their meaning? We have seen above that these representations can't just be mental images and there's a further reason why mental images don't work for the computer metaphor of the mind: mental images don't have syntax like language does. You can't create meaningful sentences by putting together a series of pictures because there are no rules for how those pictures create a holistic meaning out of the parts. For example, how could pictures represent the thought, *Leslie wants to go out in the rain but not without an umbrella* with a picture (or pictures)? How do I represent with a picture someone's desire? Or how do I represent the negation of something with only a picture? True, there are devices that we can use within pictures, such as the "no" symbol on no smoking signs. But those symbols are already not functioning purely as pictorial representations that seem to represent in virtue of their similarity. There is no pictorial similarity between the purely logical notion "not" and any picture we could draw. So whatever the words of the language of thought (that is, mental representations) are, their meaning cannot derive from a pictorial similarity to what they represent. So we need some other account. Philosophers have given many such accounts, but most of those accounts attempt to understand mental representation in terms of a causal relationship between objects in the world and representations. That is, whatever types of objects cause (or would cause) certain brain states to "light up," so to speak, are what those brain states represent. So if there's a particular brain state that lights up any time I see (or think about) a dog, then that is what those mental representations stand for. Delving into the nuances of contemporary theories of representation is beyond the scope of this chapter, but the important point is that the language of thought idea that these theories support is supposed to be compatible with physicalism as well as the computer analogy of explaining the mind. On this account, the "words" of the language of thought have original intentionality and thinking is just the manipulation of these "words" using certain syntactic rules (the "program") that are hard-wired into the brain (either innately or by learning) and which are akin to the grammar of a natural language.

There is a famous objection to the computer analogy of human thought that comes from the philosopher John Searle, who thinks that it shows that human thought and understanding cannot be reduced to the kind of thing that a computer can do. Searle's thought experiment is called **the Chinese Room**. Imagine that there is

a room with a man inside of it. What the man does is take slips of paper that are passed into the room via a slit. The slips of paper have writing on them that look like this:

与丈且两甤叆三

The room also contains a giant bookshelf with many different volumes of books. Those books are labeled something like this:

Volume 23: Patterns that begin with "与丈且两..."

When the man sees the slip of paper with the characters "与丈且两甤叆三" he goes to the bookshelf and pulls out volume 23 and looks for this exact pattern. When he finds it, he looks up the exact pattern and finds an entry that looks something like this:

When you see "与丈且两甤叆三" write "冇不互丙"

The man writes the symbols and then passes it back through the slit in the wall. From the perspective of the man in the room, this is what he does. Nothing more nothing less. The man inside the room doesn't understand what these symbols mean; they are just meaningless squiggles on a page to him. He sees the difference between the different symbols merely in terms of their shapes. However, from outside the room Chinese speakers who are writing questions on the slips of paper and passing them through the slot in the room come to believe that the Chinese room (or something inside it) understands Chinese and is thus intelligent.

The Chinese room is essentially a scenario in which a computer program passes the **Turing Test**. In paper published in 1950, Alan Turing proposed a test for how we should determine whether or not a machine can think. Basically, the test is whether or not the machine can make a human investigator believe that the machine is a human. The human investigator is able to ask the machine any questions they can think of (which Turing imagined would be conducted via types responses on a keyboard). Imagine what some of the questions might be. Here is one such potential question one might ask:

> Rotate a capital letter "D" 90 degrees counterclockwise and place it atop a capital letter "J." What kind of weather does this make you think of?

A computer that could pass the Turing Test would be able to answer questions such as this and thus would make a human investigator believe that the computer was actually another human being. Turing thought that if a machine could do this, we should count that machine as having intelligence. The Chinese Room thought experiment is supposed to challenge Turing's claim that something that can pass the Turing Test is

thereby intelligent. The essence of a computer is that of a **syntactic machine**—a machine that takes symbols as inputs, manipulates symbols in accordance with a series of rules (the program), and gives the outputs that the rules dictate. Importantly, we can understand what syntactic machines do without having to say that they interpret or understand their inputs/outputs. In fact, a syntactic machine cannot possibly understand the symbols because there's nothing there to understand. For example, in the case of modern-day computers, the symbols being processed are strings of 1s and 0s, which are physically instantiated in the CPU of a computer as a series of on/off voltages (that is, transistors). Note that a series of voltages are no more inherently meaningful than a series of different fluttering patterns of a flag waving in the wind, or a series of waves hitting a beach, or a series of footsteps on a busy New York City subway platform. They are merely physical patterns, nothing more, nothing less. What a computer does, in essence, is "reads" these inputs and gives outputs in accordance with the program. This simple theoretical (mathematical) device is called a "**Turing machine**," after Alan Turing. A calculator is an example of a simple Turing machine. In contrast, a modern day computer is an example of what is called a "**universal Turing machine**"— universal because it can run any number of different programs that will allow it to compute all kinds of different outputs. In contrast, a simple calculator is only running a couple different simple programs—ones that correspond to the different kinds of mathematical functions the calculator has (+, −, ×, ÷). The Chinese room has all the essential parts of the computer and is functioning exactly as a computer does: he "reads" these symbols and produces outputs using symbols, in accordance with what the program dictates. If the program is sufficiently well written, then the man's responses (the room's output) will be able to convince someone outside the room that the room (or something inside it) understands Chinese.

But the whole point is that the there is nothing inside the room that understands Chinese. The man in the room doesn't understand Chinese—they are just meaningless symbols to him. The written volumes don't understand Chinese either—how could they?—books don't understand things. Furthermore, Searle argues that the understanding of Chinese doesn't just magically emerge from the combination of all the parts of the Chinese room: if no one of the parts of the room has any understanding of Chinese, then neither does the whole room. Thus, the Chinese room thought experiment is supposed to be a counterexample to the Turing Test: the Chinese room passes the Turing Test but the Chinese room doesn't understand Chinese. Rather, it just acts as if it understands Chinese. Without understanding, there can be no thought. The Chinese room, impressive as it is for passing the Turing Test, lacks any understanding and therefore is not really thinking. Likewise, a computer cannot think because a computer is merely a syntactic machine that does not understand the inputs or the outputs. Rather, from the perspective of the computer, the strings of 1s and 0s are just meaningless symbols.[2] The people outside the Chinese room might ascribe thought and understanding of

---

2. We could actually retell the Chinese room thought experiment in such a way that what the man inside the room was manipulating was strings of 1s and 0s (what is called "binary code"). The point remains the same in either case: whether

Chinese to the room, but there is neither thought nor understanding involved. Likewise, at some point in the future, someone may finally create a computer program that would pass the Turing Test[3] and we might think that machine has thought and understanding, but the Chinese room is supposed to show that we would be wrong to think this. No merely syntactic machine could ever think because no merely syntactic machine could ever understand. That is the point of the Chinese room thought experiment.

We could put this point in terms of the distinction between original vs. derived intentionality: no amount of derived intentionality will ever get you original intentionality. Computers have only derived intentionality and since genuine thought requires original intentionality, it follows that computers could never think. Here is a reconstructed version of the Chinese room argument:

1. Computers are merely syntactic machines.
2. Therefore, computers lack original intentionality (from 1)
3. Thought requires original intentionality.
4. Therefore, computers cannot think (from 2-3)

How should we assess the Chinese room argument? One thing to say is that it seems to make a lot of simplifying assumptions about his Chinese room. For example, the philosopher Daniel Dennett suggests that in order to pass the Turing Test a computer would need something on the order of 100 billion lines of code. That would take the man inside the room many lifetimes to hand simulate the code in the way that we are invited to imagine. Searle thinks that these practical kinds of considerations can be dismissed—for example, we can just imagine that the man inside the room can operate faster than the speed of light. Searle thinks that these kinds of assumptions are not problematic, for why should mere speed of operation make any difference to the theoretical point he is trying to make—which is that the merely syntactic processing of a digital computer could not achieve understanding? Dennett, on the other hand, thinks that such simplifying assumptions should alert us that there is something fishy going on with the Chinese room thought experiment. If we were really, truly imagining a computer program that could pass the Turing Test, Dennett thinks, then it wouldn't sound nearly as absurd to say that the computer had thought.

There's a deeper objection to the Chinese room argument. This response is sometimes referred to as the "other minds reply." The essence of the Chinese room rebuttal of the Turing Test involves, so to speak, looking at the guts of what is going on inside of a computer. When you look at it "up close," it certainly doesn't seem like all of that syntactic processing adds up to intelligent thought. However, one can make exactly the same point

---

the program is defined over Chinese characters or strings of 1s and 0s, from the perspective of the room, none of it has any meaning and there's no understanding required in giving the appropriate outputs.

3. Nothing has yet, claims to the contrary notwithstanding.

about the human brain (something that Searle believes *is undoubtedly* capable of thought): the functioning of neurons, or even whole populations of neurons in neuronal spike trains, do not look like what we think of as intelligent thought. Far from it! But of course it doesn't follow that human brains aren't thinking! The problem is that in both cases we are looking at the wrong level of description. In order for us to be able to "see" the thought, we must be looking in the right place.

Zooming in and looking at the mechanics of the machines up close is not going to enable us to see the thought and intelligence. Rather, we have to zoom out to the level of behavior and observe the responses in their context. Thought isn't something we can see up close; rather, thought is something that we attribute to something whose behavior is sufficiently intelligent. Dennett suggests the following cartoon as a reductio ad absurdum of the Chinese room argument:

In the cartoon, Dennett imagines someone going inside the Chinese room to see what is going on inside the room. Once inside they see the man responding to the questions of Chinese speakers outside the room. The woman tells the man (perhaps someone she knows), "I didn't know you knew Chinese!" In response the man explains that he doesn't and that he is just looking up the relevant strings Chinese characters to write in response to the inputs he receives. The woman's interpretation of this is: "I see! You use your understanding of

English in order to fake understanding Chinese!" The man's response is: "What makes you think I understand English?" The joke is that the woman's evidence for thinking that the man inside the room understands English is her evidence of his spoken behavior. This is exactly the same evidence that the Chinese speakers have of the Chinese room. So if the evidence is good enough for the woman inside the room to say that the man inside the room understands Chinese, why is the evidence of the Chinese speakers outside the room any different? We can make the problem even more acute. Suppose that we were to look inside the man inside the room's brain. We would see all kinds of neural activity and then we could say, "Hey, this doesn't look like thought; it's just bunches of neurons sending chemical messages back and forth and those chemical signals have no inherent meaning." Dennett's point is that this response makes the same kind of mistake that Searle makes in supposing a computer can't think: in both cases, we are focusing on the wrong level of detail. Neither the innards of the brain nor the innards of a computer looks like there's thinking going on. Rather, thinking only emerges at the behavioral level; it only emerges when we are listening to what people are *saying* and, more generally, observing what they are *doing*. This is what is called the **other minds reply** to the Chinese room argument.

Interlude: Interpretationism and Representation

The other minds reply points us towards a radically different account of the nature of thought and representation. A common assumption in the philosophy of mind (and one that Searle also makes) is that thought (intentionality, representation) is something to be found within the inner workings of the thinking thing, whether we are talking about human minds or artificial minds. In contrast, on the account that Dennett defends, thought is not a phenomenon to be observed at the level of the inner workings of the machine. Rather, thought is something that we attribute to people in order to understand and predict their behaviors. To be sure, the brain is a complex mechanism that causes our intelligent behaviors (as well as our unintelligent ones), but to try to look inside the brain for some language-like representation system is to look in the wrong place. Representations aren't something we will find in the brain, they are just something that we attribute to certain kinds of intelligent things (paradigmatically human beings) in order to better understand those beings and predict their behaviors. This view of the nature of representation is called **interpretationism** and can be seen as a kind of **instrumentalism**. Instrumentalists about representation believe that representations aren't, in the end, real things.

Rather, they are useful fictions that we attribute in order to understand and predict certain behaviors. For example, if I am playing against the computer in a game of chess, I might *explain* the computer's behavior by attributing certain thoughts to it such as, "The computer moved the pawn in front of the king because it *thought* that I would put the king in check with my bishop and it didn't *want* to be in check." I might also attribute thoughts to the computer in order to *predict* what it will do next: "Since the computer would rather lose its pawn than its rook, it will move the pawn in front of the king rather than the rook." None

of this requires that there be internal representations inside the computer that correspond to the linguistic representations we attribute. The fundamental insight about representation, according to interpretationism, is that just as we merely *interpret computers as* having internal representations (without being committed to the idea that they actually contain those representations internally), so too we merely interpret human beings as having internal representations (without being committed to whether or not they contain those internal representations). It is useful (for the purposes of explaining behavior) to interpret humans as having internal representations, even if they don't actually have internal representations.

Interpretationist accounts of representation raise deep questions about where meaning and intentionality reside, if not in the brain, but we will not be able to broach those questions here. Suffice it to say that the disagreement between Searle and Dennett regarding Searle's Chinese room thought experiment traces back to what I would argue is the most fundamental rift within the philosophy of mind: the rift between the Cartesian view of the mind, on the one hand, and the behaviorist tradition of the mind, on the other. Searle's view of the mind, specifically his notion of "original intentionality," traces back to a Cartesian view of the mind. On this view, the mind contains something special—something that cannot be capture merely by "matter in motion" or by any kind of physical mechanism. The mind is sui generis and is set apart from the rest of nature. For Searle, meaning and understand have to issue back to an "original" mean-er or understand-er. And that understand-er cannot be a mindless mechanism (which is why Searle thinks that computers can't think). For Searle, like Descartes, thinking is reserved for a special (one might say, magical) kind of substance. Although Searle himself rejects Descartes's conclusion that the mind is nonphysical, he retains the Cartesian idea that thinking is carried out by a special, quasi-magical kind of substance. Searle thinks that this substance is the brain, an object that he thinks contains special causal powers and that cannot be replicated or copied in any other kind of physical object (for example, an artificial brain made out of metal and silicon). Dennett's behaviorist view of the mind sees the mind as nothing other than a complex physical mechanism that churns out intelligent behaviors that we then classify using a special mental vocabulary—the vocabulary of "minds," "thoughts," "representations," and "intentionality." The puzzle for Dennett's behaviorist view is: How can there be meaning and understanding without any original meaner/understander? How can there be only derived intentionality and no original intentionality?

Consciousness and the mind

Interpretationism sees the mind as a certain kind of useful fiction: we attribute representational states (thoughts) to people in virtue of their intelligent behavior and we do so in order to explain and predict their behavior. The causes of one's intelligent behavior are real, but the representational states that we attribute need not map neatly onto any particular brain states. Thus, there need not be any particular brain state that represents the content, "Brittney Spears is a washed up pop star," for example.

But there another aspect of our mental lives that seems more difficult to explain away in the way interpretationism explains away representation and intentionality. This aspect of our mind is **first-person conscious experience**. To borrow a term from Thomas Nagel, conscious experience refers to the "what it's like" of our first person experience of the world. For example, I am sitting here at my table with a blue thermos filled with coffee. The coffee has a distinctive, qualitative smell which would be difficult to describe to someone who has never smelled it before. Likewise, the blue of the thermos has a distinctive visual quality—a "what it's like"—that is different from what it's like to see blue. These experiences—the smell of the coffee, the look of the blue—are aspects of my conscious experience and they have a distinctive qualitative dimension—there is something it's like to smell coffee and to see blue. This qualitative character seems in some sense to be ineffable—that is, it would be very difficult if not impossible to convey what it is like to someone who had never smelled coffee or to someone who had never seen the color blue. Imagine someone who was colorblind. How would you explain what blue was to them? Sure, you could tell them that it was the color of the ocean, but that would not convey to them the particular quality that you (someone who is not color blind) experience when you look at a brilliant blue ocean or lake. Philosophers have coined a term that they use to refer to the qualitative aspects of our conscious experience: **qualia**. It seems that our conscious experience is real and cannot be explained away in the way that representation can. Maybe there needn't be anything similar to sentences in my brain, but how could there not be colors, smells, feels? The feeling of stubbing your toe and the feeling of an orgasm are very different feels (thank goodness), but it seems that they are both very much real things. That is, if neuroscientists were to be able to explain exactly how your brain causes you to respond to stubbing your toe, such an explanation would seem to leave something out if it neglected the feeling of the pain. From our first person perspective, our experiences seem to be the most real thing there are, so it doesn't seem that we could explain their reality away.

Physicalists need not disagree that conscious experiences are real; they would simply claim that they are ultimately just physical states of our brain. Although that might seem to be a plausible position, there are well known problems with claiming that conscious experiences are nothing other than physical states of our brain. The problem is that it does not seem that our conscious experience could just reduce to brain states—that is, to our neurons in our brain sending lots and lots of chemical messages back and forth simultaneously. The 17th century philosopher Gottfried Wilhelm Leibniz (1646-1716) was no brain scientist (that would take another 250 to develop) but he put forward a famous objection to the idea that consciousness could be reduced to any kind of mechanism (and the brain is one giant, complex mechanism). Leibniz's objection is sometimes referred to as "**Leibniz's mill**." In 1714, Leibniz wrote:

> Moreover, we must confess that *perception*, and what depends on it, is *inexplicable in terms of mechanical reasons*, that is, through shapes and motions. If we imagine that there is a machine whose structure makes it think, sense, and have perceptions, we could conceive it enlarged, keeping the same proportions, so that we could enter into it, as one enters into a mill. Assuming that, when inspecting

its interior, we will only find parts that push one another, and we will never find anything to explain a perception (*Monadology*, section 17).

Leibniz uses a famous form of argument here called **reductio ad absurdum**: He assumes for the sake of the argument that thinking is a mechanical process and then shows how that leads to the conclusion that thinking *cannot be* a mechanical process.We could put Leibniz's exact same point into the language of 21st century neuroscience: imagine that you could enlarge the size of the brain (in a sense, we can already do with the help of the tools of modern neuroscience). If we were to enter into the brain (perhaps by shrinking ourselves down) we would see all kinds of physical processes going on (billions of neurons sending chemical signals back and forth). However, to observe all of these processes would not be to observe the conscious experiences of the person whose brain we were observing. That means that conscious experiences cannot reduce to physical brain mechanics. The simple point being made is that in conscious experience there exist all kinds of qualitative properties (qualia)—red, blue, the smell of coffee, the feeling of getting your back scratched—but none of these properties would be the properties observed in observing someone's brain. All you will find on the inside is "parts that push one another" and never the properties that appear to us in first-person conscious experience.

The philosopher David Chalmers has coined a term for the problem that Leibniz was getting at. He calls it **the hard problem of consciousness** and contrasts it with **easy problems of consciousness**. The "easy" problems of mind science involve questions about how the brain carries out functions that enable certain kinds of behaviors—functions such as discriminating stimuli, integrating information, and using the information to control behavior. These problems are far from easy in any normal sense—in fact, they are some of the most difficult problems of science. Consider, for example, how speech production occurs. How is it that I decide what exactly to say in response to a criticism someone has just made of me? The physical processes involved are numerous and include the sounds waves of the person's question hitting my eardrum, those physical signals being carried to the brain, that information being integrated with the rest of my knowledge and, eventually, my motor cortex sending certain signals to my vocal chords that then produce the sounds, "I think you're misunderstanding what I mean when I said..." or whatever I end up saying. We are still a long way from understanding how this process works, but it seems like the kind of problem that can be solved by doing more of the same kinds of science that we've been doing. In short, solving easy problems involves understanding the complex causal mechanisms of the brain. In contrast, the hard problem is the problem of explaining how physical processes in the brain give rise to first- person conscious experience. The hard problem does not seem to be the kind of problem that could be solved by simply investigating in more detail the complex causal mechanism that is the brain. Rather, it seems to be a conceptual problem: how could it be that the colors, and sounds, the smells that constitute our first-person conscious experience of the world are nothing other than neurons firing electrical-chemical signals back and forth? As Leibniz pointed out over 250 years ago, the one seems to be a radically different kind of thing than the other.

In fact, it seems that a human being could have all of the functioning of normal human being and yet lack any

conscious experience. There is a term for such a being: a **philosophical zombie**. Philosophical zombies are by definition being that are functionally indistinguishable from you or I but who lack any conscious experience. If we assume that it's the functioning of the brain that causes all of our intelligent behaviors, then it isn't clear what conscious experience could possibly add to our repertoire of intelligent behaviors. Philosophical zombies can help illustrate the hard problem of consciousness since if such creatures are theoretically possible then consciousness doesn't seem to reduce to any kind of brain functioning. By hypothesis the brain of the normal human being and the brain of the philosophical zombie are identical. It's just that the latter lacks consciousness whereas the former doesn't. If this is possible then it does indeed seems to make consciousness seem like quite a mysterious thing for the physicalist.

There are two other famous thought experiments that illustrate the hard problem of consciousness: Frank Jackson's knowledge argument and Thomas Nagel's what it's like to be a bat argument.

Nagel's argument against physicalism turns on a colorful example: Could we (human beings) imagine what it would be like to be a bat? Although bats are still mammals, and thus not so different than human beings phylogenetically, their experience would seem to be radically different than ours. Bats echolocate around in the darkness, they eat bugs at night, and they sleep while hanging upside down. Human beings could try to do all these things, but even if they did, they would arguably not be experiencing these activities like a bat does. And yet it seems pretty clear that bats (being mammals) have some kind of subjective experience of the world—a "what it's like" to be a bat. The problem is that although we can figure out all kinds of physical facts about bats—how they echolocate, how they catch insects in the dark, and so on—we cannot ever know what it's like to be a bat. For example, although we could understand enough scientifically to be able to send signals to the bat that would trick it into trying to land on what it perceived as a ledge, we could not know what it's like for the bat to perceive an object as a ledge. That is, we could understand the causal mechanisms that make the bat *do what the bat does*, but that would not help us to answer the question of what it's like to *experience the world the way a bat experiences the world*. Nagel notes that it is characteristic of science to study physical facts (such as how the brain works) that can be understood in a third-person kind of way. That is, anyone with the relevant training can understand a scientific fact. If you studied the physics of echolocation and also a lot of neuroscience of bat brains, you would be able to understand how a bat does what a bat does. But this understanding would seem to bring you no closer to what it's like to be a bat—that is, to the first-person perspective of the bat. We can refer to the facts revealed in first-person conscious experience as **phenomenal facts**. Phenomenal facts are things like what it's like to see blue or smell coffee or experience sexual pleasure...or echolocate around the world in total darkness. Phenomenal facts are qualia, to use our earlier term. Nagel's point is that if the phenomenal facts of conscious experience are only accessible from a first-person perspective and scientific facts are always third-person, then it follows that phenomenal facts cannot be grasped scientifically. Here is a reconstruction of Nagel's argument:

1. The phenomenal facts presented in conscious experience are knowable only from the first-person

(subjective) perspective.

2. Physical facts can always be known from third-person (objective) perspective.
3. Nothing that is knowable only from the first person perspective could be the same as (reduce to) something that is knowable from the third-person perspective.
4. Therefore, the phenomenal facts of conscious experience are not the same as physical facts about the brain. (from 1-3)
5. Therefore, physicalism is false. (from 4)

Nagel uses an interesting analogy to explain what's wrong with physicalism—the claim that conscious states are nothing other than brain states. He imagines an ancient Greek saying that "matter is energy." It turns out that this statement is true (Einstein's famous $E = mc2$) but an ancient Greek person could not have possibly understood how it could be true. The problem is that the ancient Greek person could not have had the conceptual resources needed for being able to understand what this statements means. Nagel claims that we are in the same position today when we say something like "conscious states are brain states" is true. It might be true, we just cannot understand what that could possibly mean yet because we don't have the conceptual resources for understanding how this could be true. And the conceptual problem is what Nagel is trying to make clear in the above argument. This is another way at getting at the hard problem of consciousness.

Frank Jackson's famous knowledge argument is similar and makes a similar point. Jackson imagines a super scientist, whom he dubs "Mary," knows all the physical facts about color vision. Not only is she the world's expert on color vision, she knows all there is to know about color vision. She can explain how certain wavelengths of light strike the cones in the retina and send signals via the optic nerve to the brain. She understands how the brain interprets these signals and eventually communicates with the motor cortex that sends signals to produce speech such as, "that rose is a brilliant color of red." Mary understands all the causal processes of the brain that are connected to color vision. However, Mary understands this without ever having experienced any color. Jackson imagines that this is because she has been kept in a black and white room and has only ever had access to black and white things. So the books she reads and the things she investigates of the outside world (via a black and white monitor in her black and white room) are only ever black and white, never any other color. Now what will happen when Mary is released from the room and sees color for the first time? Suppose she is released and sees a red rose. What will she say? Jackson's claim was that Mary will be surprised because she will learn something new: she will learn what it's like to see red. But by hypothesis, Mary already knew all the physical facts of color vision. Thus, it follows that this new phenomenal fact that Mary learns (specifically, what it's like to see red) is not the same as the physical facts about the brain (which by hypothesis she already knows).

1. Mary knows all the physical facts about color vision.
2. When Mary is released from the room and sees red for the first time, she learns something new—the phenomenal fact of what it's like to see red.

3. Therefore, phenomenal facts are not physical facts. (from 1-2)

4. Therefore, physicalism is false. (from 3)

The upshot of both Nagel and Jackson's arguments is that the phenomenal facts of conscious experience—qualia—are not reducible to brain states. This is the hard problem of consciousness and it is the mind-body problem that arises in particular for physicalism. The hard problem is the reason why physicalists can't simply claim a victory over dualism by invoking Ockham's razor. Ockham's razor assumes that the two competing explanations equally explain all the facts but that one does so in a simpler way than the other. The problem is that if physicalism cannot explain the nature of consciousness—in particular, how brain states give rise to conscious experience—then there is something that physicalism cannot explain and, therefore, physicalists cannot so simply invoke Ockham's razor.

Two responses to the hard problem

We will consider two contemporary responses to the hard problem: David Chalmers's **panpsychism** and Daniel Dennett's **eliminativism**. Although both Chalmers and Dennett exist within a tradition of philosophy that privileges scientific explanation and is broadly physicalist, they have two radically different ways of addressing the hard problem. Chalmers's response accepts that consciousness is real and that solving the hard problem will require quite a radical change in how we conceptualize the world. On the other hand, Dennett's response attempts to argue that the hard problem isn't really a problem because it rests on a misunderstanding of the nature of consciousness. For Dennett, consciousness is a kind of illusion and isn't ultimately real, whereas for Chalmers consciousness is the most real thing we know. The disagreement between these two philosophers returns as, again, to the most fundamental divide within the philosophy of mind: that between Cartesians, on the one hand, and behaviorists, on the other.

To understand Chalmers's response to the hard problem, we must first understand what he means by a "basic entity." A basic entity is one that science posits but that cannot be further analyzed in terms of any other kind of entity. Can you think of what kinds of entities would fit this description? Or which *science* you would look to in order to find basic entities? If you're thinking physics, then you're correct. Think of an atom. Originally, atoms were thought of as the most basic building blocks of the universe; the term "atom" literally means "uncuttable" (from the Greek "a" = *not* + "tomos" = *cut*). So atoms were originally thought of as basic entities because there was nothing smaller. As we now know, this turned out to be incorrect because there were even smaller particles such as electrons, protons, quarks, and so on. But eventually physics will discover those basic entities that cannot be reduced to anything further. Mental states are not typically thought of as basic entities because they are studied by a higher order science—psychology and neuroscience. So mental states, such as my

perception of the red rose, are not basic entities. For example, brain states are ultimately analyzable in terms of brain chemistry and chemistry, in turn, is ultimately analyzable in terms of physics (not that anyone would care to carry out that analysis!). But Chalmers's radical claim is that consciousness is a basic entity. That is, the qualia—what it's like to see red, smell coffee, and so on—that constitute our first-person conscious experience of the world cannot be further analyzed in terms of any other thing. They are what they are and nothing else. This doesn't mean that our conscious experiences don't correlate with the existence of certain brain states, according to Chalmers. Perhaps my experience of the smell of coffee correlates with a certain kind of brain state. But Chalmers's point is that that correlation is basic; the coffee smell qualia are not the same thing as the brain state with which they might be correlated. Rather, the brain state and the conscious experience are just two radically different things that happen to be correlated. Whereas brain states reduce to further, more basic, entities, conscious states don't. As Chalmers sees it, the science of consciousness should proceed by studying these correlations. We might discover all kinds of things about the nature of consciousness by treating the science of consciousness as irreducibly correlational. Chalmers suggests as an orienting principle the idea that consciousness emerges as a function of the "informational integration" of an organism (including artificially intelligent "organisms"). What is informational integration? In short, **informational integration** refers to the complexity of the organism's control mechanism—its "brain." Simple organisms have very few inputs from the environment and their "brains" manipulate that information in fairly simple ways. Take an ant, for example. We pretty much understand exactly how ants work and as far as animals go, they are pretty simple. We can basically already duplicate the level of intelligence of an ant with machines that we can build. So an informational integration of an ant's brain is pretty low. A thermostat has some level of informational integration, too. For example, it takes in information about the ambient temperature of a room and then sends a signal to either turn the furnace on or off depending on the temperature reading. That is a very simple behavior and the informational integration inside the "brain" of a thermostat is very simple. Chalmers's idea is that complex consciousness like our emerges when the informational integration is high—that is, when we are dealing with a very complex brain. The less complex the brain, the less rich the conscious experience. Here is a law that Chalmers suggests could orient the scientific study of consciousness:
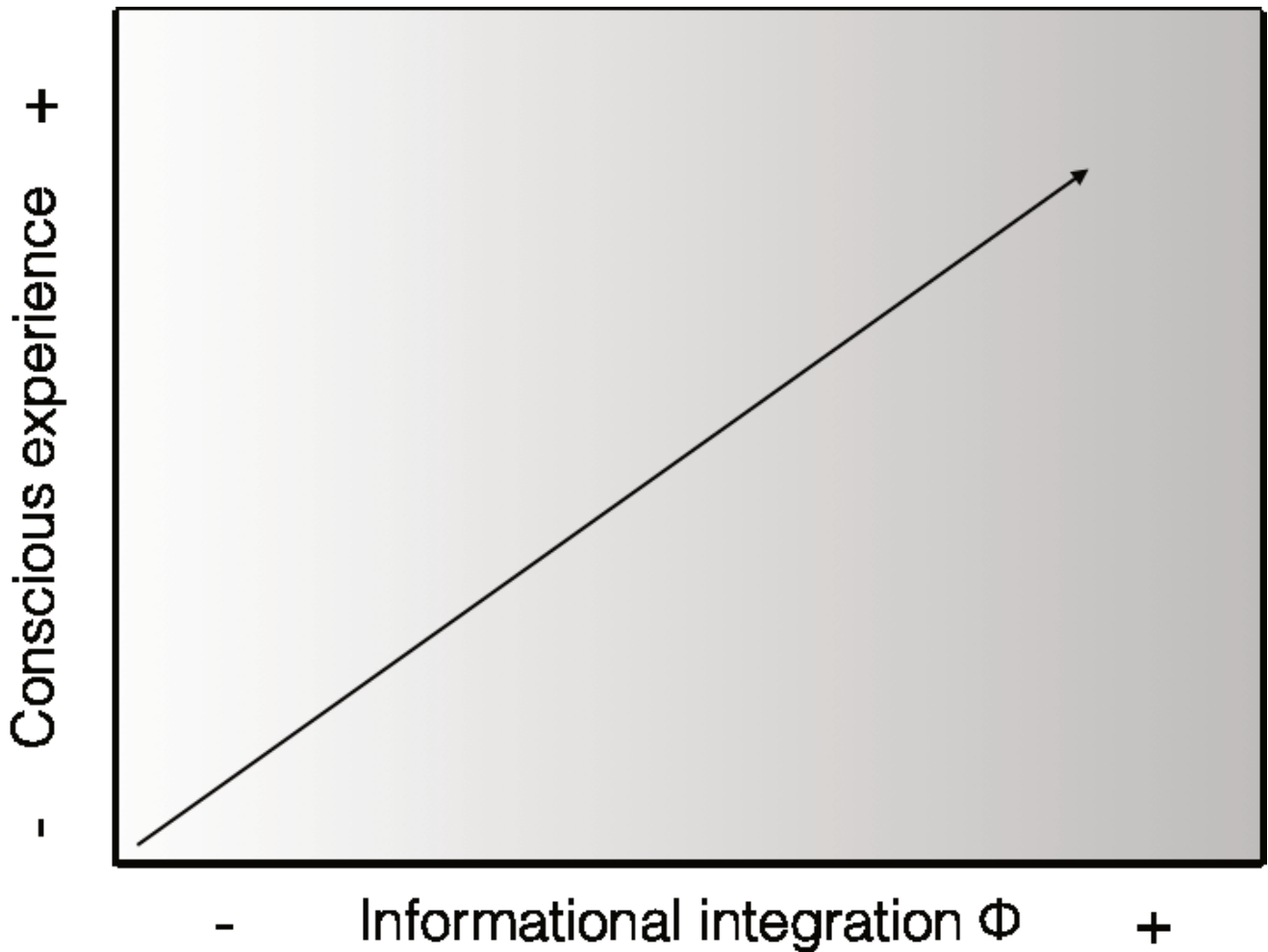
This graph just says that as informational integration increases, so does the complexity of the associated conscious experience. Again, the conscious experience doesn't reduce to informational integration, since that would only run headlong into the hard problem—a problem that Chalmers thinks is unsolvable.

The graph also says something else. As drawn, it looks like even information processing systems whose informational integration is low (for example, a thermostat or tree) also has some non-negligible level of conscious experience. That is a strange idea; no one really thinks that a thermostat is conscious and the idea that plants might have some level of conscious experience will seem strange to most. This idea is sometimes referred to as panpsychism ("pan" = all, "psyche" = mind)—there is "mind" distributed throughout everything in the world. Panpsychism is a radical departure from traditional Western views of the mind, which sees minds as the purview of animals and, on some views, of *human beings alone*. Chalmers's panpsychism still draws a line between objects that process information (things like thermostats, sunflowers, and so on) and those that don't (such as rocks), but it is still quite a radical departure from traditional Western views. It is not, however, a radical departure from all sorts of older, prescientific and indigenous views of the natural world according to which everything in the natural world, including plants and streams, as possessing some sort of spirit—a mind

of some sort. In any case, Chalmers thinks that there are other interpretations of his view that don't require the move to panpsychism. For example, perhaps conscious experience only emerges once information processing reaches a certain level of complexity. This interpretation would be more consistent with traditional Western views of the mind in the sense that one could specify that only organisms with a very complex information processing system, such as the human brain, possess conscious experience. (Graphically, based on the above graph, this would mean the lowest level of conscious experience wouldn't start until much higher up the y-axis.)

Daniel Dennett's response to the hard problem fundamentally differs from Chalmers's. Whereas Chalmers posits qualia as real aspects of our conscious experience, Dennett attempts to deny that qualia exist. Rather, Dennett thinks that consciousness is a kind of illusion foisted upon us by our brain. Dennett's perennial favorite example to begin to illustrate the illusion of consciousness concerns our visual field. From our perspective, the world presented to us visually looks to be unified in color and not possessing any "holes." However, we know that this is not actually the case. The cones in the retina do not exist on the periphery and, as a result, you are not actually seeing colors in the objects at the periphery of your visual field. (You can test this by having someone hold up a new object on one side of your visual field and moving it back and forth until you are able to see the motion. Then try to guess the color of the object. Although you'll be able to see the object's motion, you won't have a clue as to its color, if you do it correctly.) Although it seems to us as if there is a visual field that is wholly colored, it isn't really that way. This is the illusion of consciousness that Dennett is trying to get us to acknowledge; things are not really as they appear. There's another aspect of this illusion of our visual field: our blind spot. The location where the optic nerve exits the retina does not convey any visual information since there are no photoreceptors; this is known as the blind spot. There are all kinds of *illustrations to reveal your blind spot*. However, the important point that Dennett wants to make is that from our first-person conscious experience it never appears that there is any gap in our picture of the world. And yet we know that there is. This again is an illustration of what Dennett means by the illusion of conscious experience. Dennett does more than simply give fun examples that illustrate the strangeness of consciousness; he has also famously attacked the idea that there are qualia. Recall that qualia are the purely qualitative aspects of our conscious experiences—for example, the smell of coffee, the feeling of a painful sunburn (as opposed to the pain of a headache), or the feeling of an orgasm. Qualia are what are supposed to create problems for the physicalist since it doesn't seem that that purely qualitative feels could be nothing more than the buzzing of neurons in the brain. Since qualia are what create the trouble for the physicalism and since Dennett is a physicalist, one can understand why Dennett targets qualia and tries to convince us that they don't exist.

If you're going to argue against something's existence, the best way to do that is first precisely define what it is you are trying to deny. Then you argue that as defined such things cannot exist. This is exactly what

Dennett does with qualia.[4] He defines qualia as the qualitative aspects of our first-person conscious experience that are a) irreducibly first-person (meaning that they are inaccessible to third-person, objective investigation) and b) intrinsic properties of one's conscious experience (meaning that they are what they are independent of anything else). Dennett argues that these two properties (irreducibly first person and intrinsic) are in tension with each other—that is, there can't be an entity which possesses both of these properties. But since both of these properties are part of the definition of qualia, it follows that qualia can't exist—they're like a square circle.

Change blindness is a widely studied phenomenon in cognitive psychology. Some of the demonstrations of it are quite amazing and have made it into the popular media many times over the last couple of decades. One of the most popular research paradigms to study change blindness is called the flicker paradigm. In the flicked paradigm, two images that are the same with the exception of some fairly obvious difference are exchanged in a fairly rapid succession, with a "mask" (black or white screen) between them. What is surprising is that it is very difficult to see even fairly large differences between the two images. So let's suppose that you are viewing these flickering images and trying to figure out what the difference between them is but that you haven't yet figured it out yet. As Dennett notes, there are of course all kinds of changes going on in your brain as these images flicker. For example, the photoreceptors are changing with the changing images. In the case of a patch of color that is changing between the two images, the cones in your retina are conveying different information for each image. Dennett asks: "Before you noticed the changing color, were your color qualia changing for that region?" The problem is that any way you answer this question spells defeat for the defender of qualia because either they have to give up (a) their irreducible subjectiveness or their intrinsicness (b). So suppose the answer to Dennett's question is that your qualia *are* changing. In that case, you do not have any special or privileged access to your qualia, in which case they aren't irreducibly subjective, since subjective phenomena are by definition something we alone have access to. So it seems that the defender of qualia should reject this answer. Then suppose, on the other hand, that your qualia *aren't* changing. In that case, your qualia can't change unless you notice them changing. But that makes it looks like qualia aren't really intrinsic, after all since their reality is constituted by whether you notice them or not. And "noticings" are relational properties, not intrinsic properties. Furthermore, Dennett notes that if the existence of qualia depend on one's ability to notice or report them, then even philosophical zombies would have qualia, since noticings/reports are behavioral or functional properties and philosophical zombies would have these by definition. So it seems that the qualia defender should reject this answer as well. But in that case, there's no plausible answer that the qualia defender can give to Dennett's question. Dennett's argument has the form of a classic **dilemma**, as illustrated below:

---

4. Daniel Dennett, Sweet Dreams: Philosophical Obstacles to a Science of Consciousness. MIT Press. 2006.

Yes → Not irreducibly subjective → **qualia don't exist**

Did your qualia change before you noticed?

No → Not intrinsic → **qualia don't exist**
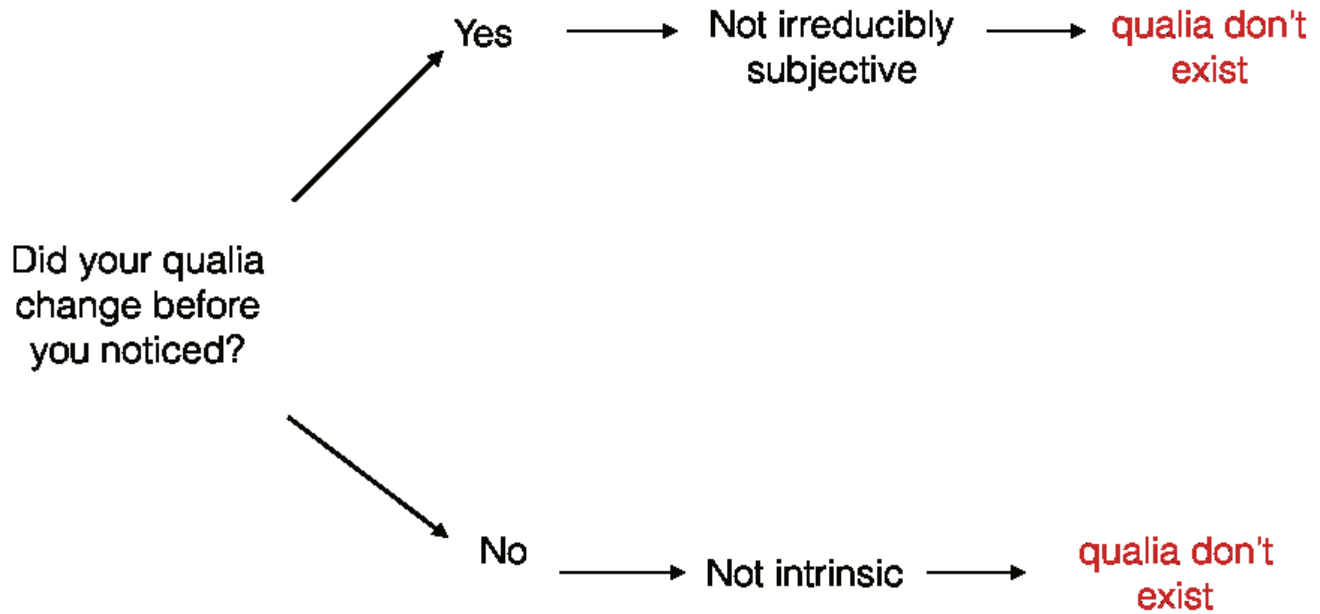
Dennett thinks that the reason there is no good answer to the question is that the concept of qualia is actually deeply confused and should be rejected. But if we reject the existence qualia it seems that we reject the existence of the thing that was supposed to have caused problems for physicalism in the first place. Qualia are a kind of illusion and once we realize this, the only task will be to explain why we have this illusion rather than trying to accommodate them in our metaphysical view of the world. The latter is Chalmers's approach whereas the former is Dennett's.

Study questions

1. True or false: One popular way of thinking about how the mind works is by analogy with how a computer works: the brain is a complex syntactic engine that uses its own kind of language—a language that has original intentionality.
2. True or false: One good way of explaining how the mind understands things is to posit a little man inside the head that does the understanding.
3. True or false: The mind-body problem is the same, exact problem for both physicalism and dualism.
4. True or false: John Searle agrees with Alan Turing that the relevant test for whether a machine can think is the test of whether or not the machine behaves in a way that convinces us it is intelligent.
5. True or false: One good reply to the Chinese Room argument is just to note that we have exactly the same behavioral evidence that other people have minds as we would of a machine that passed the Turing Test.
6. True or false: According to interpretationism, mental representations are things we attribute to others in order to help us predict and explain their behaviors, and therefore it follows that mental representations

must be real.

7. True or false: This chapter considers two different aspects of our mental lives: mental representation (or intentionality) and consciousness. But the two really reduce to the exact same philosophical problem of mind.

8. True or false: The hard problem is the problem of understanding how the brain causes intelligent behavior.

9. True of false: The knowledge argument is an argument against physicalism.

10. True or false: Dennett's solution to the hard problem turns out to be the same as Chalmers's solution.

For deeper thought

1. How does the hard problem differ from the easy problems of brain science?

2. If the Turing Test isn't the best test for determining whether a machine is thinking, can you think of a better test?

3. According to physics, nothing in the world is really red in the way we perceive it. Rather, redness is just a certain wavelength of light that our senses interpret in a particular way (some other creature's sensory system might interpret that same physical phenomenon in a very different way). By the same token, redness does not exist in the brain: if you are seeing red then I cannot also see the red by looking at your brain. In this case, where is the redness if it isn't in the world and it also isn't in the brain? And does this prove that redness is not a physical thing, thus vindicating dualism? Why or why not?

4. Could someone be in pain and yet not know it? If so, how would we be able to tell they were in pain? If not, then aren't pain qualia real? And so wouldn't that prove that qualia are real (if pain is)?

5. According to Chalmers's view, is it theoretically possible for a machine to be conscious? Why or why not?

CHAPTER 4: PERSONAL IDENTITY

PERSONAL IDENTITY

Matthew Van Cleave

Introduction: the persistence question

Imagine a wooden ship called "TS" that was built 100 years ago and that has been in continual use up until today. As time has passed, parts of TS have decayed and have needed to be replaced. Eventually, none of TS's parts are original; all of them have been replaced. Is TS the same boat as it was 100 years ago?[1] On the one hand, it seems that it cannot be since the physical object has undergone a radical change—it is literally a totally different physical object. One the other hand, we are inclined to still call it "TS" since it has been in continual use since it was first built and it has always been called "TS." These two answers suggest two very different criteria about the identity conditions of objects. The first answer suggests that an object is the same object only if there is some **intrinsic property** of the object that stays the same through time (and that if there are no such properties then it follows that it is no longer the same ship as it was at the start). The second answer suggests that an object is the same object as long as it has some extrinsic property that remains such same. Being called the by the same name (or perceived by people to be the same object) over time is a good example of an **extrinsic property** since it is not something that the object itself contains but, rather, something that is "put onto" the objection from outside. On this criterion, TS is the same ship since people have always called it "TS."

The question about the identity of the boat represents a very general **metaphysical** question about the identity of physical objects: what are the criteria by which objects remain numerically identical? **Numerical identity** is a term that philosophers use to describe an object being the very same object. It is contrasted with **qualitative identity** which simply means that an object has all the same properties or qualities. An object is only numerically identical to itself but can be qualitatively identical to other objects. For example, imagine two identical red Corvettes. The two red Corvettes are qualitatively identical (since if you closed your eyes could not tell whether one has been substituted for the other), but they are not numerically identical (because they are two different cars). The **persistence question of identity** is the question about the conditions under

---

1. This thought experiment is quite old—going back to Ancient Greece at least—but the first record of it we have in writing is from Plutarch. He writes: "The ship wherein Theseus and the youth of Athens returned had thirty oars, and was preserved by the Athenians down even to the time of Demetrius Phalereus, for they took away the old planks as they decayed, putting in new and stronger timber in their place, insomuch that this ship became a standing example among the philosophers, for the logical question of things that grow; one side holding that the ship remained the same, and the other contending that it was not the same."

which an object remains numerically identical. For example, if the ship were disassembled and the wood used to build a house, it seems clear that it would no longer be the same object (even if its physical material were the same) since it is no longer a boat but a house. Likewise, if the red Corvette were melted down into metal and then fashioned into a boat, it would no longer be the same object.

Although we can ask the persistence question about objects in general, one kind of thing whose identity conditions philosophers have been very interested in is persons. The **persistence question of personal identity** is the question about the conditions under which a particular human being remains numerically the same.[2] The philosophical problem is that although it seems that that a person remains numerically identical throughout their life, it is very difficult to give a satisfying answer to what accounts for this numerical identity. More specifically, the problem is that although we are inclined to think that there is some intrinsic property that makes a person numerically the same throughout their life, it is very difficult to come up with what this intrinsic property is. Perhaps there is no intrinsic property; perhaps the only thing that makes me me is that others have always called me "Matt Van Cleave." However, that answer also seems to lack the depth that we think our personal identity has. After all, it seems that there is a clear sense in which I would still be me even if all of a sudden people started calling me by a different name. Suppose I was kidnapped and taken to a different place where no one knew me but where I looked exactly like someone called Todd Quiring. The fact that everyone now calls me "Todd Quiring" doesn't seem to change who I am in any deep way.

In this chapter we will consider a few different answers to the persistence question. John Locke's psychological continuity account and Bernard Williams's bodily continuity account both attempt to find an intrinsic property that grounds our personal identity. In contrast, Daniel Dennett's short story "Where am I?" raises the question of whether there is any intrinsic property that grounds personal identity. In philosophy, this idea is famously linked to David Hume, but it also connects with the Buddhist concept of *anattā* ("no self"). Whereas the persistence question is a metaphysical question, in the last part of the chapter we will consider what I call the **psychological question of personal identity**: Where does my *sense that* that I am the same person throughout my life come from? Notice that even if it were to turn out that metaphysically there is no self that persists throughout our whole lives, it still makes sense to ask this psychological question about our *sense of* our identity. That is, even if there is no deep sense in which I am the same person I was 20 years ago, there can still be an account of what gives me the *sense that* I am (if indeed I have that sense).[3]

---

2. I say "human being" rather than "person" here because as we will see below, different answers to the persistence question may or may not focus on the identity of persons. "Human being" is a broader term which leaves open the question of personhood. For example, I have always been a human being, but (depending on one's view) I may not have always possessed personhood (for example, when I was an infant) and I may continue to persist as a human being even if I am no longer a person (for example, if I lose all of my brain functioning).

3. Some philosophers have claimed not to have the sense that they are the same person from one moment to the next. David

Personal identity as an unchanging, intrinsic property

Before considering Locke's view of personal identity, let's consider one idea about what kind of thing would answer the persistence question. It is commonly thought our personal identity should consist in some one thing that is unchanging and intrinsic. For example, within certain religious traditions, the idea of an immaterial soul is often seen as the thing that makes us the same person. Although our bodies, our beliefs, and our values may change radically throughout our lifetimes, our soul is thought to stay the same. However, if it is possible for our souls to leave our bodies and inhabit other bodies (such as in reincarnation) then it seems problematic for souls to ground *my* personal identity. If one soul can inhabit many different persons, then it cannot be the soul that distinguishes one person from another. In addition, positing a soul raises certain kinds of **epistemological** questions: How do we know that souls exist? Even if souls exist, how do we know they don't themselves undergo change? How do I know that my soul is the same from one moment to the next? For these reasons and others, philosophers have tended to look elsewhere for an answer the persistence question of personal identity.

Is there some other thing about me that always stays the same? What about my body? Like the example of the ship, my body naturally regenerates all of its physical matter certain number of years (how long depends on which organs/cells we are talking about). Although neurons in our cerebral cortex do not regenerate, the connections between the neurons in our brain are continually changing (this is what happens when learning takes place). In any case, it seems implausible to maintain that some is some particular neuron or set of neurons in the cerebral cortex that account for why I am the same. What about DNA? The DNA within our bodily cells stays largely the same (except when uncorrected errors in replication occur, which is rare). But consider identical twins (which have the same DNA). Clearly identical twins are two different persons. But if DNA were the basis of personal identity, there would be only one person. How about fingerprints—don't those stay the same throughout my lifetime? Yes, they do—unless I were to lose my hands. But in that case, am I really now a different person? It seems not. The other problem with DNA and fingerprints is that they don't seem to have much to do with who we are as *persons*. Many philosophers since Locke have viewed the concept of a person as a **forensic concept**, meaning a concept that is essentially tied to our notions of moral and legal responsibility. Persons are individuals have have moral and legal rights and can be held responsible for their actions. The problem with DNA and fingerprints is that neither one has much to do with our personhood and agency and for this reason wouldn't be able to explain much about *personal* identity. Thus it seems that

---

Hume famously claimed (A Treatise of Human Nature, book 1, part IV, section VI). that we did not have any awareness of a persisting self. And the Oxford philosopher Derek Parfit came to see one's sense of identity is a kind of illusion and that it doesn't really matter that we do not persist as persons—in fact, that it can actually have salutary effects on one's ethical stance towards the world. More on this idea in the last part of the chapter.

there is no one physical thing that persists throughout our lifetime that could ground our personal identity. For this reason, philosophical accounts of personal identity have eschewed the idea that there is some *one thing* that stays same throughout my life. Instead, philosophers have tended to claim that what grounds identity is a kind of continuity between aspects of ourselves at different times.

Personal identity as psychological continuity

John Locke (1632-1704) claimed that I remain the same person through time because I am conscious of being the same person through time and that "as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person."[4] Locke's criterion is commonly interpreted as a **memory criterion of personal identity**: personal identity consists in one's ability to remember things about their own past from a first-person perspective. For example, since I can remember things about being 20 years old (or even 10 years old) it follows that I am now numerically identical to the person as I was then. The fact that these memories are from the first-person perspective is crucial. It isn't just that I can remember certain events taking place (since perhaps some of my friends remember those same events). Rather, it is that I remember those events *from my first-person perspective*. For Locke, what makes a person the same person must be essentially tied to what persons are and for Locke "person" is a **forensic concept**, meaning a concept that is essentially tied to the concept of **agency** and thus to our notions of moral and legal responsibility. Persons are individuals who have moral and legal rights and who can be held responsible for their actions. (For this reason, Locke would reject DNA and fingerprints as grounding *personal* identity since neither one has much to do with the capacities that underlie our lives as agents.)

The memory criterion has a broad appeal since it seems plausible that much of our sense of ourselves is due to our memories of our past. It does seem that without the ability to remember anything about my past (as in cases of severe amnesia or dissociative fugue states) I would not be the same person. Furthermore, the memory criterion of personal identity carries with it a criterion for moral responsibility: we cannot be held responsible for things of which we have no memory. The memory criterion is also appealing because it accounts for the commonly held idea that our minds are central to who we are as persons (since memory is a psychological attribute). Finally, since psychological properties are intrinsic properties of persons, Locke's memory criterion accounts for the commonly held belief that our personal identities are real and deep. So far, so good. However, the memory criterion is not without its problems.

Suppose that Bob commits a heinous murder when he is 25 years old and is never caught for it. 50 years later he is finally apprehended but now Bob has late stage Alzheimer's disease and cannot remember anything about his

---

4. John Locke, *An Essay Concerning Human Understanding*, book 2, chapter 27.

earlier life. What would Locke's memory criterion say about whether Bob should be punished for the earlier murder? The memory criterion implies that Bob is literally not the same person now as he was then. If we assume that it is fair only to punish the person who committed the crime, it would not be fair to punish Bob today since he is a different person (according to the memory criterion). Think of it this way: it is no more fair to punish Bob for the murder than it is to punish his friend Bill, who had nothing to do with the murder. On Locke's memory criterion, Bob at 75 years old is no more the same person than Bob at 25 years old than his friend Bill is. Neither one has the right kind of memory connection to Bob at 25 years old. In fact, since no existing person has the right kind of memory connection to 25 year old Bob, it follows that there is no existing person that committed the crime (and thus no one can be punished). Locke anticipated this problem and accepts that if Bob truly cannot remember his earlier life, then it is not fair to punish him (since he is not the same person who committed the murder). However, Locke also thinks that we should still punish people who have committed crimes even if they claim they can't remember it since if we didn't, then anyone found guilty of a crime could claim they didn't remember the crime as a way of exculpating themselves. So Locke thinks that we have to hold people accountable. If we were omniscient like god, then we would know when someone really couldn't remember (and thus wasn't the same person) and when they were just lying to escape punishment. But since we aren't, we are stuck with an inferior kind of justice that may sometimes punish those who don't deserve it.

Another problem that arises for Locke's memory criterion is what happens when we are in a state in which we cannot remember anything from our earlier lives— for example, when I am unconscious, such as I am when I sleep. Someone who is asleep does not have the ability to remember earlier events in one's life and thus it seems to follow that when a person is asleep they are not the same person as when they are awake. Indeed, if personhood requires the ability to remember at all, then it seems that the unconscious person is not really a person at all. But this seems all wrong. It does not seem that when I am sleeping I cease to be numerically identical to the person I am when I am awake. One way of trying to solve this problem would be to broaden the memory criterion. For example, you could say that it is still possible for the sleeping person to remember earlier events in their life, whereas it isn't possible for the severe Alzheimer's patient to remember such events. If the sleeping person were simply woken up, they would be able to remember their earlier life. But there is no simple thing that could be done to restore the Alzheimer's patient's memory in this way.[5]

One of the most famous objections to the memory criterion turns on the idea that numerical identity is a **transitive relationship**. When we say that x is numerically identical to y, we mean that x and y are one and

---

5. Some philosophers would put this point in terms of "possible worlds." You could say that there is a "nearby possible world" in which, say, Bill, who is merely asleep, could remember earlier episodes of his life (all you have to do is wake Bill up). In contrast, there is no nearby possible world in which Bob, who as severe Alzheimer's, could be made to remember earlier episodes of his life.

the same object. That is, if x is identical to y and y is identical to z, must be true that x is identical to z. Once we understand this, we can see that there is a problem with the memory criterion. Suppose that Roses at age 50 can remember things about Roses at age 25 and that Roses at age 75 can remember things about Roses at age 50. Does it follow that Roses at age 75 can remember things about Roses at age 25? It should be clear that it doesn't follow and this shows that memory is not in general a transitive relationship. Now here is the problem. According to the memory criterion two persons are identical if the later one can first- personally remember things that happened to the earlier one. So Roses at 25 and Roses at 50 are identical. Likewise, Roses at 50 and Roses at 75 are identical. But as we saw above, numerical identity is transitive, so it must follow that Roses at 25 and Roses at 75 are numerically identical. But as we have just seen, it doesn't follow because memory is not transitive in the way numerical identity is. Therefore, memory cannot be the same as numerical identity.

Therefore, the memory criterion cannot be correct.[6] How might Locke respond to this objection? On the one hand, Locke seems happy to admit that you have not always been the same person throughout your life (since you aren't when you're sleeping) and that you may become a wholly different person (as in the case of Bob, above). So Locke might be content to just allow that the lifetime of Roses consists of more than one different person. On the other hand, if we are looking for a sense of personal identity that lasts throughout a person's lifetime, then in some cases (for example, cases where there are memory failures) we will have to say that there are multiple different persons that have existed in the course of one lifetime.

Philosophers continue to debate Locke's memory criterion, but many philosophers influenced by Locke have tended to move to a broader **psychological continuity account of personal identity**. On these accounts, what makes us a person numerically identical through time is that there are causal connections between our psychological states. For example, at age 10 I had a certain set of beliefs, desires, and memories (all different kinds of psychological states). At age 11 I probably had a similar set of psychological states, although of course new ones had been added and some changed. But what makes me at 11 the same as me at age 10 is that the

---

6. This objection to the memory criterion of personal identity was originally raised by Thomas Reid in his *Essays on the Intellectual Powers of Man* (essay 3, chapter 6), which was published in 1785. Reid says: *"Suppose a brave officer to have been flogged when a boy at school, for robbing an orchard, to have taken a standard from the enemy in his first campaign, and to have been made a general inadvanced life : suppose also, which must be admitted to be possible, that, when he took the standard, he was conscious of his having been flogged at school, and that when made a general he was conscious of his taking the standard, but had absolutely lost the consciousness of his flogging. These things being supposed, it follows, from Mr. Locke's doctrine, that he who was flogged at school is the same person who took the standard, and that he who took the standard is the same person who was made a general. Whence it follows, if there be any truth in logic, that the general is the same person with him who was flogged at school. But the general's consciousness does not reach so far back as his flogging therefore, according to Mr. Locke's doctrine, he is not the person who was flogged. Therefore the general is, and at the same time is not, the same person with him who was flogged at school."*

psychological states I had at age 10 were causally involved in bringing about the psychological states I had at age 11. It is these causal relationships between my psychological states that tie me back to these older versions of myself and thus account for my persistence through time.

Psychological continuity accounts of personal identity retain the Lockean identification of our persisting identities with our psychological states, but they reject the Lockean idea that these must be consciously accessible mental states. So, in a sense, *I* don't have to be able to maintain the persistence of my identity through time (for example, by consciously connecting myself now to myself ten years ago). Rather, my identity is maintained by the causal relationships between my psychological states. The result of making this relationship causal is that my personal identity can persist despite my lack of memory. This solved several of the problems that were raised for the memory criterion. For example, if I am asleep or awake my persisting beliefs, desires, and values do not change. When I am asleep, I still believe that George Washington was the first U.S. president, that water contains hydrogen atoms, that I had a piñata at my tenth birthday party, and liberal democracies are in theory the least bad form of government. What makes me the same person while I am asleep is that my psychological states while asleep are caused by my previous psychological states while I was awake. It doesn't matter, as it did for Locke's memory criterion, that I cannot consciously access these psychological states while asleep. Likewise, making the continuity between our psychological states causal rather than conscious solves the transitivity problem explained above. It doesn't matter that Roses at 75 cannot remember anything about Roses at 25. Rather, what matters is that Roses's psychological states at 75 can be traced back in causal line to Roses's psychological states at 25. Of course, her psychological states may have changed radically in that time, but that doesn't matter. In this way, psychological continuity accounts allow for the possibility of persistence despite radical change. And in some sense this is exactly what our common sense intuitions suggest about the persistence of our identities through time: although we can change radically, nonetheless, there is still a sense in which we remain the same.

But how radical can the change be and the individual still persist? The most difficult problem for psychological continuity accounts is called the fission problem. Suppose we perform a hemispherectomy (splitting the brain in two at the corpus callosum and then removing one half) on "Lefty" and transplant the right half of his brain into another person (call him "Righty"). The psychological continuity view would seem to imply that Lefty and Righty are the same person since both are psychologically continuous with Lefty. But it seems there are now two people whereas the psychological continuity view implies that there is only one person. This is the fission problem. One thing the psychological continuity theorist could say in response to the fission problem is that since there is a radical, exogenous causal disruption of Lefty's psychological states ("half" of them are removed and put into another person), that Lefty ceases to exist since the causal continuity of his psychological states is disrupted. In this case, one person ceases to exist and two different ones, Lefty2 and Righty come into existence. This seems a sensible way of responding to bizarre thought experiments like the one above. After all, it isn't intuitively obvious that Lefty continues to exist after such a radical procedure. On the other hand,

if we think that Lefty remains numerically identical even after the hemispherectomy, then the psychological continuity view would imply that Righty is also Lefty and that seems to be an absurdity.

Personal identity as bodily continuity

The bodily continuity account of personal identity, as I am using that term here, contrasts with psychological continuity accounts in that it claims that what accounts for our persistence through time is our bodily continuity, not our psychological continuity. A famous argument for the bodily continuity account comes from the philosopher Bernard Williams.[7] Williams has us imagine the following sci-fi thought experiment. Suppose that you are told that you will be painfully tortured tomorrow but that before you are a certain procedure will be performed on you: all of your psychological states (memories, beliefs, desires, and so on) will be wiped clean. That is, you will become complete amnesiac, remembering nothing of who you are or were. You will become essentially a blank slate. Psychological continuity accounts would seem to imply that in this state you have ceased to exist. Thus, the person that will be tortured tomorrow will not be you. However, it seems that this would be cold comfort to the person facing this prospect. That is, it seems we would still now fear the torture that will occur tomorrow, *even if none of our psychological states remain when we are undergoing the torture*. To Williams, the fact that we still fear the torture implies that we do not identity ourselves with our psychological states, but rather with our bodies. How else could you explain why we would fear the torture? It seems that the psychological continuity account of our identity would say that we shouldn't fear the torture since the person being tortured tomorrow will not be us. Be since we do fear the torture, this implies that we *do* identify ourselves with that individual who will be tortured tomorrow.

Williams argues that this is so even if we imagine switching all of our psychological states with another person—that is, all of their psychological states are transferred to our brain/body and all of ours, to their brain/body. Again, the psychological continuity view would seem to imply that the personal identities of these two bodies have switched: the A-body person takes on B's identity and the B-body person takes on A's identity. Figure 1 below is a representation of this situation.

---

7. Williams, Bernard, 1970, "The Self and the Future," *The Philosophical Review*, 79: 161-180. Reprinted numerous times/places.

Figure 1
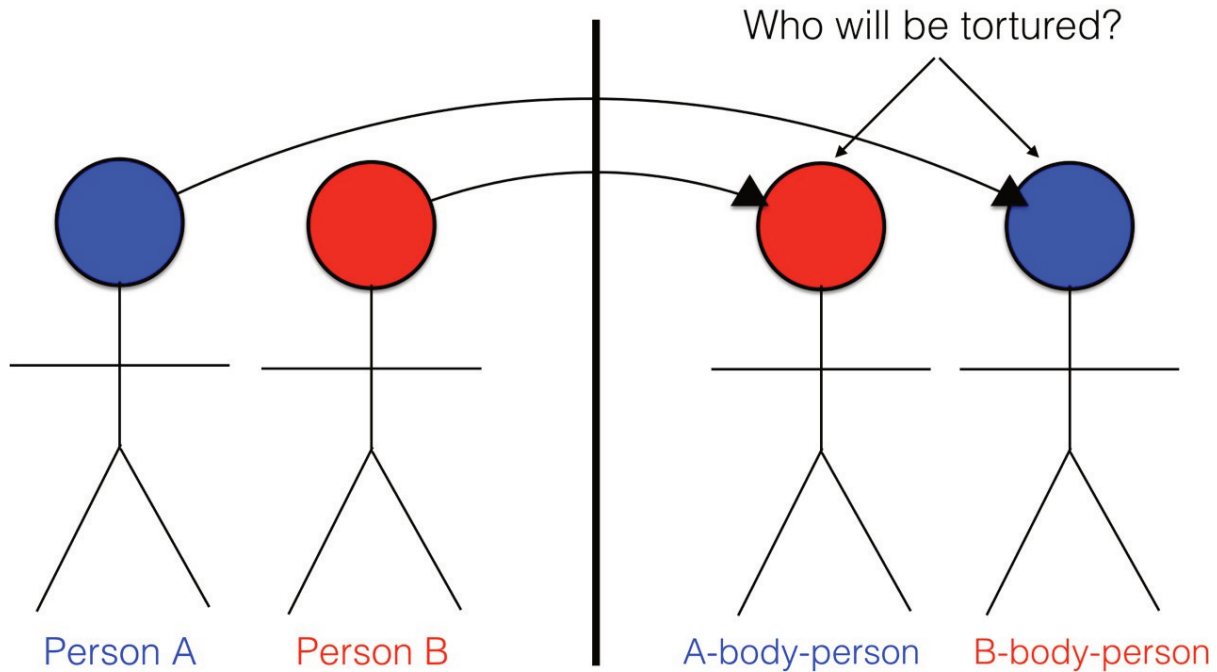
Suppose that you are A in the above scenario and suppose that you are reasoning purely in terms of self-interest (that is, you want to avoid being tortured). Who would you say should be tortured tomorrow: the A-body person or the B-body person? Williams think that we would say the B-body person should be tortured, *even if the B-body person inherits all of our psychological states*! Since psychological continuity accounts (including the Lockean memory criterion) imply that person A now exists in the B-body person, then such accounts imply that we should say that the A-body person should be tortured. The fact that we would say that the B-body person should be tortured is supposed to clearly refute psychological continuity theories, according to Williams (or at least to raise a serious problem for them). Again, what we seem to be doing in a case like this is identifying ourselves not with the continuity of our psychological states, but rather with the continuity of our bodies.

If you aren't already convinced that if you were A, you would identity with the A- body person after the transfer of psychological states (and with the prospect of the impending doom of torture applied to the A-body person), Williams walks through a series of thought experiments, starting with the above amnesia scenario. Since he thinks we would clearly still fear torture (now) even if we were to be given total amnesia before the torture took place, he uses this to show that there is no difference between the amnesia scenario and the psychological transfer scenario, described above. Willliams walks us through the following permutations of the amnesiac scenario in order to arrive at the transfer scenario, claiming that none of the changes should make a difference to our answer of who should be tortured:

- Total amnesia
- Total amnesia + totally new memories created
- Total amnesia + totally new memories *that come from some other existing person,* call them "B-person"
- Total amnesia + totally new memories *that come from some other existing person* + your memories (that have been wiped clean) are given to that other B-person)

This last scenario is just the same transfer scenario described above. Williams claims that if you said that the B-body person should be tortured in the amnesia scenario then you should logically say the same thing in the transfer scenario, since there are no relevant changes in any of the scenarios through which the amnesia scenario is transformed into the transfer scenario.

One puzzling thing about Williams's article is that he seems to equate transfer scenarios like the one above (where all psychological states are transferred) to *brain transfer* scenarios and it is not at all obvious that this is so.[8] Although it seems to me correct that we should identify with our bodies rather than our mental states in the *psychological state* transfer scenario, we should not identify our bodies in a *brain* transfer scenario. If you are A in the scenario and your brain will be transferred to B's body, then it seems to me that you should choose the A-body person to be tortured. In this case, you won't feel it because the seat of consciousness that underlies our experiences (whether of pleasure or pain) has now shifted, with the transfer of my brain, to B's body. Therefore, you should choose the A-body person to be tortured. If this is right, then it seems like Williams's scenario doesn't really show that we identify with our bodies, as such. Rather, it shows that we identify with our brains—specifically with our first- person conscious experience that we think goes where our brain goes. (As we will see presently, Dennett's sci-fi story will challenge this intuition.) The remains behind when we transfer all of our psychological states is the seat of conscious experience—that which is capable of experiencing pleasure and pain. The reason we would choose the B-body person to be tortured is that although they have all of our psychological states, they do not have our seat of consciousness—our first-person conscious experience. That remains behind with our brain, even if all of our psychological states are extracted from it and replace with B's psychological states. In contrast, in the brain transfer scenario, not only all of our psychological states are being transferred, so also is our seat of conscious experience itself. But since the "seat of conscious experience" is as much a *psychological* property as it is a *physical* property (the brain), Williams's argument doesn't really show that we identify our persistence with some aspect of our bodies (brains) rather than some aspect of our minds (consciousness).[9]

---

8. Williams (1970), p. 162.

9. The question of the relationship between our brains and our consciousness is thorny philosophical question. Some philosophers think that consciousness reduces to brain processes while other thinks that consciousness cannot be strictly reduced to brain processes. For more on this topic, see the chapter in this textbook on the mind-body problem.

Within contemporary philosophical debates about the persistence question, the view that aligns most closely with bodily accounts of personal identity is a view called **animalism**.[10] Contrary to Lockean accounts of the persistence of our personal identity in terms of our psychological properties, animalists claim that what accounts for our persistence through time is our animal properties. Yes, we are *thinking* animals, but we are first and foremost animals. According to animalists, our "animal properties" are the ones that account for our persistence through time as the same animal. For example, this could the biological properties that maintain the life of the organism over time—they are what keep the organism from dying and becoming, through decomposition, something else. This means that what accounts for the persistence of individual human beings through time is the same kind of thing that accounts for the persistent of individual dogs or cats or trees through time. A crucial argument that animalists make against Lockean accounts is that only animalism can make sense of the idea that I persist through time from birth to death. As we have seen, accounts which ground our personal identity in our psychological properties have to admit that my personal identity stops when there is a large enough discontinuity in my psychological states. Animalists admit that although I may cease to be a person under certain conditions (for example, if I become a complete amnesiac), I remain an animal and thus *I* continue to persist through time. Insofar as we think that the essence of who we are is extremely robust and persists through the kinds of radical changes that we can undergo throughout our lifetimes, animalism is an attractive alternative to Lockean answers to the persistence question.

Another argument that animalists have made is called the **animal ancestors argument**. Suppose, contrary to what animalism claims, we are not essentially animals. If we are not animals, then our parents are not animals either, nor are our grandparents, and so on. But this entails something that is false according to evolutionary theory: that nothing in our ancestry is an animal. But since we know this to be false, it follow that our original assumption must be false. That is, it follows that we must be, in essence, animals. This argument uses a famous form of argument called a *reductio ad absurdum* (this is Latin for "reduce to absurdity). The argument proceeds by assuming the falsity of a claim P that we is true and then show how if we assume the falsity of that claim P, it leads to an absurdity (strictly, a **contradiction**). But since absurdities like contradictions cannot be true, if follows that our original assumption (that P is false) must *itself* be false, which means that P is true.

One interesting objection to animalism is the case of dicephalic twins. The problem is that it looks there are *two individuals* in the case of dicephalic twins. However, it seems pretty clear that there is only *one animal*. Dicephalic twins thus constitute a **counterexample** to animalism's claim that our identity consists in our animal properties.[11]

---

10. See the Stanford Encyclopedia of Philosophy entry here: https://plato.stanford.edu/entries/animalism/
11. See McMahan, Jeff, 2002, The Ethics of Killing: Problems at the Margins of Life, Oxford: Oxford University Press, p. 32.

Dennett's "Where Am I?"

Daniel Dennett's, "Where Am I," is a mind-bending sci-fi short story to read in order to problematize different accounts of personal identity.[12] Arguably, problematizing the concept of personal identity, as well as different accounts of it, is exactly what Dennett was trying to do in writing the story.[13] Regardless of Dennett's goal, what I will try to do in this section is use the story to show that there is no one way of accounting for our personal identity that can withstand the radical ways in which individual human beings might persist through time. If Dennett is correct, then it is possible that I (my identity) could be spread over many different places at one time (omnipresence!) or even that there could be two individuals that share my numerically identity! I highly recommend that you simply read the story for yourself before reading my own synopsis of it.

"Where Am I?" begins with Dennett signing up for an expedition in which he must disarm a highly radioactive object underground. Because it is feared that the radioactivity will destroy his brain, an operation is performed in which his brain is removed and placed into a vat of liquid, keeping it vital. In addition, electrodes are implanted both in the neurons in his severed brain and in the brainstem in his body so that the brain can still communicate with the body electronically, at a distance. If you doubt this could be done, just consider that the brain's neurons are sending many, many simple messages that could be, in theory, relayed through other mediums, like radio waves. As Dennett puts it, the doctors describe the operation procedure as essentially stretching his neurons so that they can communicate with his body from a long distance (wirelessly, by electronic signals like radio waves).

Whilst on the underground expedition, something goes wrong and Dennett's body is crushed in an accident. Before the accident it seems to Dennett, from his first-person perspective, that he is underground. However, once his body is destroyed, his point of view reverts immediately back to lab, where his brain is in a vat of liquid—miles and miles away from where he just was underground before the accident. At this point, Dennett has no body, just a brain. This would be a frightening position to be in, if you think about it. There is no sensory input, no way of communicating with the outside world.[14] While in this state of extreme sensory deprivation, scientists figure out how to stream Dennett's favorite classical music (Brahms) straight into his auditory nerve of his brain (thus bypassing the normal way of hearing via the mechanics of our ears).

Eventually, the scientists find a new body for Dennett and connect his original brain to the new body in the

---

12. This was originally published in Dennett's, *Brainstorms* (Bradford Books, 1978) and has been reprinted many times in different places.

13. That this was one of the things he was attempting to do in this piece was in fact confirmed to me by Dennett in personal correspondence (Snapchat—just kidding, email).

14. Aficionados of the heavy metal band Metallica will recognize this kind of situation as one portrayed in the song, "One."

aforementioned way (via electrodes in the new body's brainstem). The newly reimbodied Dennett emerges, now able to communicate with the rest of the world again. Immediately Dennett's point of view switches from inside the vat (where his brain is) to outside the vat, where his new body is. He walks back into the lab to the applause of all the scientists working on the project and stares at his brain inside the vat of liquid. Earlier, scientists had shown him a switch which turned on/off the transmitter that sent the brain signals electronically to/from the body. Earlier, when he had turned that switch off, his body had collapsed to the floor (since it was no longer connected to the brain). Just for kicks, Dennett now tries this again and miraculously nothing happens. His conscious experience is exactly the same; his point of view is exactly the same; nothing seems to change at all.

Puzzled by this, he asks the scientists who explain that they had actually "cloned" his brain on a computer so that there was an exact replica of his brain that was running the exact same input/output signals to his body. What the new switch did was simply change which thing was connected to his body: the original brain or the clone of his brain on the computer. The putative reason that scientists had done this was just in case Dennett's brain was destroyed, Dennett could still live on via his "brain" that the computer had cloned. As Dennett flips the switch back and forth, he can tell no difference at all in terms of his point of view and the nature of his conscious experience. It is as if nothing had happened at all.

The last part of the story concerns another upsetting circumstance (both emotionally for the character "Dennett" in the story and philosophically for Dennett, the author of the story). Dennett's original brain and the computer brain come out of sync with each other—the inputs and outputs no longer match—so that when the switch is thrown to switch what is controlling Dennett's body, a whole new persona emerges (in this case, from weeks of inability to control the body). In essence, there are now two people controlling Dennett's (new) body, but only one can control the body at a time. They both receive sensory inputs to the body, but only one can send motor signals to move the body. If you were the person who couldn't send motor signals to reach the body, you can imagine how distressing this would be.

Below is a diagram which shows each different stage of the story.
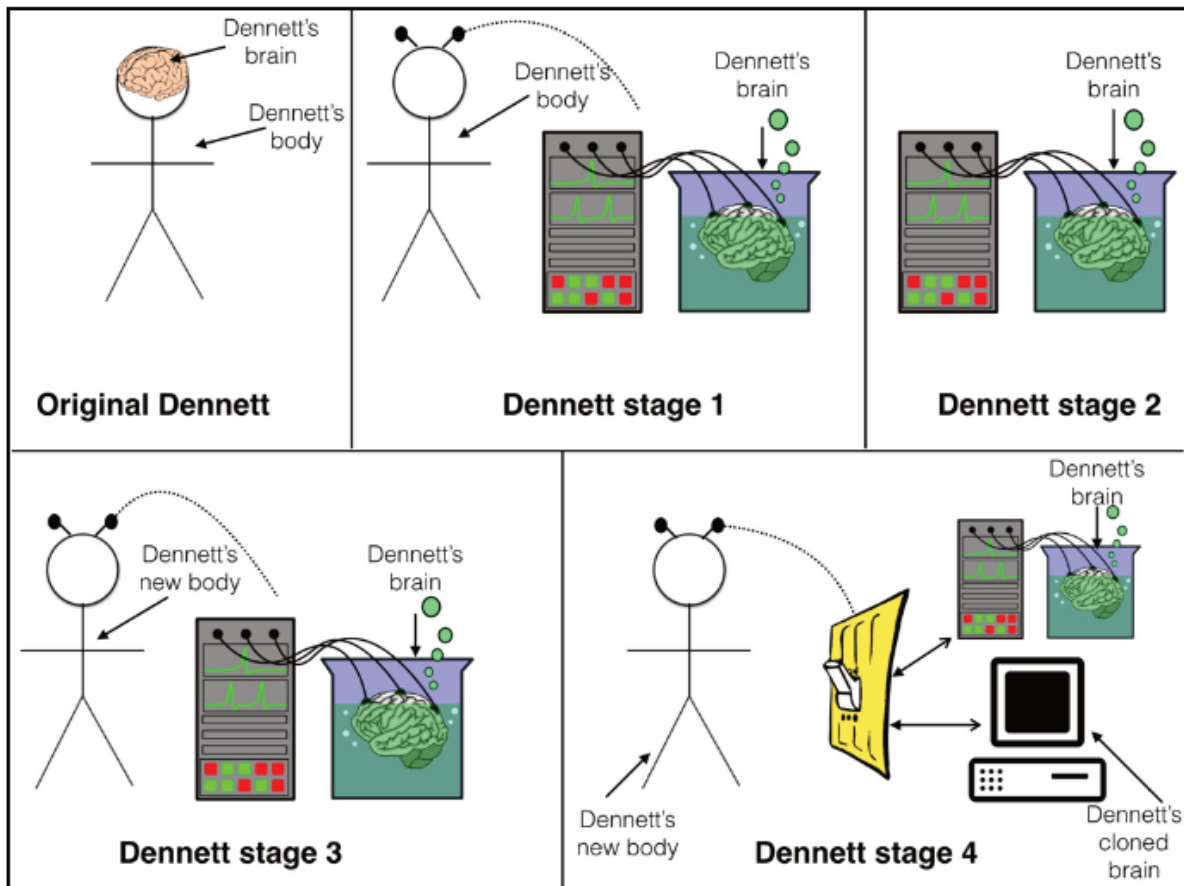
Notice what the story seems to entail. Those who say that bodily continuity accounts for our persistence would seem to be incorrect, since Dennett seems to himself in the story to persist even though his original body has been destroyed and replaced by someone else's body. On the other hand, those who would see our persisting identities as tied to our brains also seem to be mistaken, since we can imagine a further stage of the story (before the computer and brain came out of sync) in which Dennett's original brain (sitting in the vat of liquid in the lab) was destroyed in a different accident and yet Dennett's point of view remained the same. So one's identity is not bodily continuity nor is it, more specifically, the continuity of the brain. This seems to leave open the psychological continuity accounts of personal identity. However, consider one of the paradoxical implications that this story raises for psychological continuity accounts. In the last stage of the story, before the computer and Dennett's brain have gotten out of sync, it appears that there are *two, numerically identical Dennetts*, since the information (memories, beliefs, desires) on the clone of Dennett's brain on the computer is identical to Dennett's brain. (This is just an instance of the fission problem, explained above.)But there cannot be two, numerically identical objects. Therefore, psychological continuity accounts cannot be correct.

One might take the upshot of Dennett's story to be that there are no good answers to the persistence questions. That is, perhaps I do not really persist through time at all. The idea that I—my "self"—does not persist through time is an old idea that also arises in older philosophical traditions, both in the East and West. The Buddhist

concept of *anattā* ("no self") suggests the impermanence and everchanging nature of our personal identities.[15] In the West, David Hume famously claimed that:

> For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch myself at any time without a perception, and never can observe anything but the perception. When my perceptions are removed for any time, as by sound sleep; so long am I insensible of myself, and may truly be said not to exist. And were all my perceptions removed by death, and could I neither think, nor feel, nor see, nor love, nor hate after the dissolution of my body, I should be entirely annihilated, nor do I conceive what is farther requisite to make me a perfect non-entity. If anyone, upon serious and unprejudiced reflection thinks he has a different notion of himself, I must confess I can reason no longer with him. All I can allow him is, that he may be in the right as well as I, and that we are essentially different in this particular. He may, perhaps, perceive something simple and continued, which he calls himself; though I am certain there is no such principle in me. But setting aside some metaphysicians of this kind, I may venture to affirm of the rest of mankind, that they are nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement.[16]

Perhaps Hume and the Buddhists are correct. After all, we have seen that there are difficult problems that arise for the different answers to the persistence question. But even if there *is* no persisting self, it is nevertheless true that many people have *the sense* that there is. A separate question concerns how to best explain where our *sense* that we persist through time comes from. In the last couple of decades, an interdisciplinary view has become very influential: the narrative conception of the self. In the next section we will consider this view.

The narrative conception of the self

Why does it feel to me that there is a persisting self? According to the narrative conception of the self, what unifies our lives—that is, what makes it seem that me at 10 years old is the same as me at 43 years old—is the stories we tell about ourselves that connect the different parts of our lives. For example, I, Matthew Van Cleave, am the person who grew up in a small town, became a runner in high school and college, played in a rock band in college, and then after college decided to become a philosopher and so went back to graduate

---

15. One excellent representation of the doctrine of anatta comes from the Buddhist text, The Questions of King Milinda (Book 2, chapter 1) in which the Buddhist sage, Nāgasena, explains to Milinda that just as a chariot is nothing in additions to all its parts, so the self is nothing in addition to all of its parts.
16. David Hume, 1739, *A Treatise of Human Nature* (book 1, part IV, section VI).

school to earn a PhD in Philosophy. Furthermore, narratives have a certain kind of structure to them: there is beginning, middle and end. Importantly, the end of the narrative is what enables us to make sense of and assess who we are at any point within the narrative. The end of the narrative is what provides guidance throughout the narrative: what counts a good or bad decision is relative to the end goal/purpose of the life whose narrative is being narrated. For example, if the end of my story concerns being a successful philosophy professor and having a profound impact on my students and colleagues, then this end is what my life is working towards and the parts of my life constitute my self insofar as they are a part of this narrative structure. According to the narrative view of the self, who we are is just the story that we tell about ourselves. Notice that the story we tell about the events is different than the events themselves. The narratives we construct of our lives are *post-hoc* (Latin for "after this") in the sense that we construct them after the events have already happened. But as mentioned above, a narrative can also exert a *normative influence* over what we choose to do in the future. For example, I might choose not to take a much higher paying job in the business world because it doesn't cohere with my narrative of myself as an influential and respected philosopher. Or I might decide to return the credit card I found on the sidewalk instead of trying to use it because using it would be inconsistent with my narrative of myself as an honest person.

The narrative conception of self is compatible with the idea that there is nothing metaphysically deep about the self. The narrative that I construct is just one of many different possible narratives that could be constructed from the same raw material of my life—my experiences, memories, interests, commitments, and so on. Whereas philosophers have traditionally tried to respond to the persistence problem by searching for something *within* those raw materials that unifies our self and explains our persistence, the narrative view sees the unity as something imposed from the outside—from a narrator. Thus, like the example of the ship in the beginning of this chapter, the narrative view views one's identity as an extrinsic property (like the mere fact that the ship was always called "TS") rather than as an intrinsic property. Perhaps a good analogy would be currency. Consider what makes a 100 dollar bill valuable. It certainly isn't the paper an ink which constitutes the bill. Rather, it is the complex set of rules and institutions outside of the 100 dollar bill that gives it value. If you were to destroy that system (for example, if the government were overturned and all of the institutions like the Federal Reserve that bestow value on the 100 dollar bill ceased to exist), the 100 dollar bill would no longer have its value. It would simply be a piece of paper with some ink on it. Likewise, according to the narrative conception of the self, the fact that there is a persisting self depends on our ability to impose a story on our lives. If we lost that ability, there would be no persisting self although there would still be experiences, memories and thoughts. For the narrative view, the persisting self is an abstract entity that we construct from the raw materials of our lives and which, once constructed, exerts guidance on our lives.

One problem with the narrative view is that it defines the self too narrowly. If the existence of a persisting self requires a narrative arc with a beginning, middle, and end, then many individuals' lives will lack this structure, for example, because their lives ended early. Such lives would seem lack any persisting self, if the narrative view is correct. Moreover, there are many people who have a sense of themselves as persisting through time although

they don't really have any grand narrative of their life. For this reason and others, some philosophers, such as Elisabeth Camp, have suggested thinking about our constructed identities not in the narrative, beginning-middle-end way, but rather thinking of ourselves as characters in an unfolding story whose end we don't know yet. The main difference is that whereas the narrative conception holds one's interpretation of their life hostage to the end of the story, the character conception doesn't. Instead, what guides our interpretation of our lives is a "particular nexus of dispositions, memories, interests, and commitments."[17] The interpretation that we create can also be constrained by information we glean about ourselves from other people, such as our friends.

Study questions

1. True or false: The persistence question asks whether or not there exists a momentary awareness of ourselves as agents.
2. True or false: There are two main types of answer to the persistence question.
3. True or false: John Locke's view of personal identity is a kind of psychological continuity view.
4. True or false: According to Locke, if our DNA always stayed the same, this would be a good answer to the persistence question.
5. True or false: DNA could serve as a good criterion of numerical identity.
6. True or false: One of the most difficult problems for the psychological continuity account is the fission problem.
7. True or false: The fact that we should fear torture to our body even if all of our psychological states have been transferred to a different body is supposed to provide support to the bodily continuity account of personal identity.
8. True or false: According to Locke, "person" is a forensic term.
9. True or false: Animalism is a type of bodily continuity account of personal identity.
10. True or false: Dicephalic twins pose a problem for animalism.
11. True or false: Dennett's science fiction story suggests that a person can continue to persist despite having neither a brain nor a body.
12. True or false: Narrative accounts of the self are compatible with the idea that there is no persisting self.

For deeper thought

1. What does it mean to say that "identity is a transitive relationship"? Explain how Locke's memory criterion conflicts with the idea that identity is transitive.
2. Whereas Lockean accounts are interested in accounting for how persons persist, animalists think that we

---

17. Elisabeth Camp, "Wordsworth's Prelude, Poetic Autobiography, and Narrative Constructions of the Self."

should be explaining how organisms persist. Think of an example where the organism continues to persist while the person doesn't.

3. Explain the sense which the narrative account of the self makes our identities external to us. How does this different from traditional accounts of the persistence of the self?

4. How is sleep a problem for Locke's memory criterion?

5. Suppose that we say that the identity conditions of a self are simply the object that traces a continuous spatio-temporal path. What makes me the same person as I was when I was 10 years old is that the body of my 10 year old self and the body of my 43 year old self has traces a continuous, uninterrupted spatiotemporal path. Would psychological continuity theorists accept this account? Why or why not?

6. Suppose that in Williams's transfer scenarios we transfer our whole brain from one body to another. Does this change your answer of who will be tortured? Why or why not?

7. Why might Williams's transfer scenarios really prove that we identify with our bodily properties rather than with our psychological properties?

CHAPTER 5: THE PROBLEM OF FREE WILL AND DETERMINISM

THE PROBLEM OF FREE WILL AND DETERMINISM

Matthew Van Cleave

> "You say: I am not free. But I have raised and lowered my arm. Everyone understands that this illogical answer is an irrefutable proof of freedom."
>
> -Leo Tolstoy
>
> "Man can do what he wills but he cannot will what he wills."
>
> -Arthur Shopenhauer
>
> "None are more hopelessly enslaved than those who falsely believe they are free."
>
> -Johann Wolfgang von Goethe

The term "freedom" is used in many contexts, from legal, to moral, to psychological, to social, to political, to theological. The founders of the United States often extolled the virtues of "liberty" and "freedom," as well as cautioned us about how difficult they were to maintain. But what do these terms mean, exactly? What does it mean to claim that humans are (or are not) free? Almost anyone living in a liberal democracy today would affirm that freedom is a good thing, but they almost certainly do not all agree on what freedom is. With a concept as slippery as that of free will, it is not surprising that there is often disagreement. Thus, it will be important to be very clear on what precisely we are talking about when we are either affirming or denying that humans have free will. There is an important general point here that extends beyond the issue of free will: *when debating whether or not x exists, we must first be clear on defining x, otherwise we will end up simply talking past each other*. **The philosophical problem of free will and determinism** is the problem of whether or not free will exists in light of determinism. Thus, it is crucial to be clear in defining what we mean by "free will" and "determinism." As we will see, these turn out to be difficult and contested philosophical questions. In this chapter we will consider these different positions and some of the arguments for, as well as objections to, them.

Let's begin with an example. Consider the 1998 movie, *The Truman Show*. In that movie the main character, Truman Burbank (played by Jim Carrey), is the star of a reality television show. However, he doesn't know that he is. He believes he is just an ordinary person living in an ordinary neighborhood, but in fact this

neighborhood is an elaborate set of a television show in which all of his friends and acquaintances are just actors. His every moment is being filmed and broadcast to a whole world of fans that he doesn't know exists and almost every detail of his life has been carefully orchestrated and controlled by the producers of the show. For example, Truman's little town is surrounded by a lake, but since he has been conditioned to believe (falsely) that he had a traumatic boating accident in which his father died, he never has the desire to leave the small little town and venture out into the larger world (at least at first). So consider the life of Truman as described above. Is he free or not? On the one hand, he gets to do pretty much everything he wants to do and he is pretty happy. Truman doesn't look like he's being coerced in any explicit way and if you asked him if he was, he would almost certain reply that he wasn't being coerced and that he was in charge of his life. That is, he would say that he was free (at least to the same extent that the rest of us would). These points all seem to suggest that he is free. For example, when Truman decides that he would rather not take a boat ride out to explore the wider world (which initially is his decision), he is doing what he *wants* to do. His action isn't coerced and does not feel coerced to him. In contrast, if someone holds a gun to my head and tells me "your wallet or your life!" then my action of giving him my wallet is definitely coerced and feels so.

On the other hand, it seems clear the Truman's life is being manipulated and controlled in a way that undermines his agency and thus his freedom. It seems clear that Truman is not the master of his fate in the way that he thinks he is. As Goethe says in the epigraph at the beginning of this chapter, there's a sense in which people like Truman are those who are most helplessly enslaved, since Truman is subject to a massive illusion that he has no reason to suspect. In contrast, someone who knows she is a slave (such as slaves in the antebellum South in the United States) at least retains the autonomy of knowing that she is being controlled. Truman seems to be in the situation of being enslaved and not knowing it and it seems harder for such a person to escape that reality because they do not have any desire to (since they don't know they are being manipulated and controlled).

As the *Truman Show* example illustrates, it seems there can be reasonable disagreement about whether or not Truman is free. On the one hand, there's a sense in which he is free because he does what he wants and doesn't feel manipulated. On the other hand, there's a sense in which he isn't free because what he wants to do is being manipulated by forces outside of his control (namely, the producers of the show). An even better example of this kind of thing comes from Aldous Huxley's classic dystopia, *Brave New World*. In the society that Huxley envisions, everyone does what they want and no one is ever unhappy. So far this sounds utopic rather than dystopic. What makes it dystopic is the fact that this state of affairs is achieved by genetic and behavioral conditioning in a way that seems to remove any choice. The citizens of the Brave New World do what they want, yes, but they seems to have absolutely no control over what they want in the first place. Rather, their desires are essentially implanted in them by a process of conditioning long before they are old enough to understand what is going on. The citizens of *Brave New World* do what they want, but they have no control over what the want in the first place. In that sense, they are like robots: they only have the desires that are chosen for them by the architects of the society.

So are people free as long as they are doing what they want to—that is, choosing the act according to their strongest desires? If so, then notice that the citizens of *Brave New World* would count as free, as would Truman from *The Truman Show*, since these are both cases of individuals who are acting on their strongest desires. The problem is that those desires are not desires those individuals have chosen. It feels like the individuals in those scenarios are being manipulated in a way that we believe we aren't. Perhaps true freedom requires more than just that one does what one most wants to do. Perhaps true freedom requires a genuine choice. But what is a genuine choice beyond doing what one most wants to do?

Philosophers are generally of two main camps concerning the question of what free will is. **Compatibilists** believe that free will requires only that we are doing what we want to do in a way that isn't coerced—in short, free actions are voluntary actions. **Incompatibilists**, motivated by examples like the above where our desires are themselves manipulated, believe that free will requires a genuine choice and they claim that a choice is genuine *if and only if, were we given the choice to make again, we could have chosen otherwise*. I can perhaps best crystalize the difference between these two positions by moving to a theological example. Suppose that there is a god who created the universe, including humans, and who controls everything that goes on in the universe, including what humans do. But suppose that god does this not my directly coercing us to do things that we don't want to do but, rather, by implanting the desire in us to do what god wants us to do. Thus human beings, by doing what the want to do, would actually be doing what god wanted them to do. According to the compatibilist, humans in this scenario would be free since they would be doing what they want to do. According to the incompatibilist, however, humans in this scenario would not be free because given the desire that god had implanted in them, they would always end up doing the same thing if given the decision to make (assuming that desires deterministically cause behaviors). If you don't like the theological example, consider a sci-fi example which has the same exact structure. Suppose there is an eccentric neuroscientist who has figured out how to wire your brain with a mechanism by which he can implant desire into you.

Suppose that the neuroscientist implants in you the desire to start collecting stamps and you do so. However, you know none of this (the surgery to implant the device was done while you were sleeping and you are none the wiser).

From your perspective, one day you find yourself with the desire to start collecting stamps. It feels to you as though this was something you chose and were not coerced to do. However, the reality is that given this desire that the neuroscientist implanted in you, you could not have chosen *not* to have started collecting stamps (that is, you were necessitated to start collecting stamps, given the desire). Again, in this scenario the compatibilist would say that your choice to start collecting stamps was free (since it was something you wanted to do and did not feel coerced to you), but the incompatibilist would say that your choice was not free since given the implantation of the desire, you could not have chosen otherwise.

We have not quite yet gotten to the nub of the philosophical problem of free will and determinism because

we have not yet talked about determinism and the problem it is supposed to present for free will. What is determinism?

Determinism is the doctrine that every cause is itself the effect of a prior cause. More precisely, if an event (E) is determined, then there are prior conditions (C) which are sufficient for the occurrence of E. That means that if C occurs, then E *has to* occur. Determinism is simply the claim that every event in the universe is determined. Determinism is assumed in the natural sciences such as physics, chemistry and biology (with the exception of quantum physics for reasons I won't explain here). Science always assumes that any particular event has some law-like explanation—that is, underlying any particular cause is some set of law- like regularities. We might not know what the laws are, but the whole assumption of the natural sciences is that there are such laws, even if we don't currently know what they are. It is this assumption that leads scientists to search for causes and patterns in the world, as opposed to just saying that everything is random. Where determinism starts to become contentious is when we move into the human sciences, such as psychology, sociology, and economics. To illustrate why this is contentious, consider the famous example of Laplace's demon that comes from Pierre-Simon Laplace in 1814:

We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes.

Laplace's point is that if determinism were true, then everything that every happened in the universe, including every human action ever undertaken, had to have happened. Of course humans, being limited in knowledge, could never predict everything that would happen from here out, but some being that was unlimited in intelligence could do exactly that. Pause for a moment to consider what this means. If determinism is true, then Laplace's demon would have been able to predict from the point of the big bang, that you would be reading these words on this page at this exact point of time. Or that you had what you had for breakfast this morning. Or any other fact in the universe. This seems hard to believe, since it seems like some things that happen in the universe didn't have to happen. Certain human actions seem to be the paradigm case of such events. If I ate an omelet for breakfast this morning, that may be a fact but it seems strange to think that this fact was necessitated as soon as the big bang occurred. Human actions seem to have a kind of independence from web of deterministic web of causes and effects in a way that, say, billiard balls don't.

Given that the cue ball his the 8 ball with a specific velocity, at a certain angle, and taking into effect the coefficient of friction of the felt on the pool table, the exact location of the 8 ball is, so to speak, already determined before it ends up there. But human behavior doesn't seem to be like the behavior of the 8 ball in this way, which is why some people think that the human sciences are importantly different than the

natural sciences. Whether or not the human sciences are also deterministic is an issue that helps distinguish the different philosophical positions one can take on free will, as we will see presently. But the important point to see right now is that determinism is a doctrine that applies to all causes, including human actions. Thus, if some particular brain state is what ultimately caused my action and that brain state itself was caused by a prior brain state, and so on, then my action had to occur given those earlier prior events. And that entails that *I couldn't have chosen to act otherwise, given that those earlier events took place*. That means that the incompatibilist position on free will cannot be correct if determinism is true. Recall that incompatibilism requires that a choice is free only if one could have chosen differently, given all the same initial conditions. But if determinism is true, then human actions are no different than the 8 ball: given what has come before, the current event had to happen. Thus, if this morning I cooked an omelet, then my "choice" to make that omelet could not have been otherwise. Given the complex web of prior influences on my behavior, my making that omelet was determined. It *had to* occur.

Of course, it feels to us, when contemplating our own futures, that there are many different possible ways our lives might go—many possible choices to be made. But if determinism is true, then this is an illusion. In reality, there is only one way that things could go, it's just that we can't see what that is because of our limited knowledge. Consider the figure below. Each junction in the figure below represents a decision I make and let's suppose that some (much larger) decision tree like this could represent all of the possible ways my life could go. At any point in time, when contemplating what to do, it seems that I can conceive of my life going many different possible ways. Suppose that A represents one series of choices and B another. Suppose, further, that A represents what I actually do (looking backwards over my life from the future). Although from this point in time it seems that I could also have made the series of choices represented in B, if determinism is true then this is false. That is, if A is what ends up happening, then *A is the only thing that ever could have happened*. If it hasn't yet hit you how determinism conflicts with our sense of our own possibilities in life, think about that for a second.
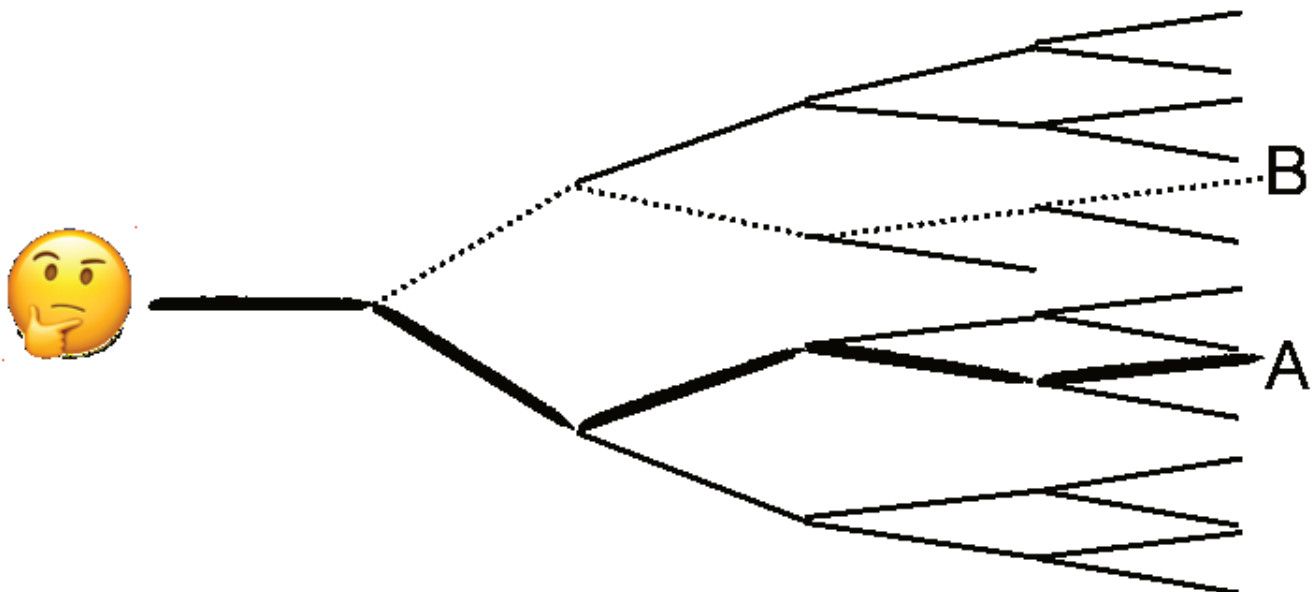
As the foregoing I hope makes clear, the incompatibilist definition of free will is incompatibile with determinism (that's why it's called "incompatibilist"). But that leaves open the question of which one is true. To say that free will and determinism are logically incompatible is just to say that they cannot both be true, meaning that one or the other must be false. But which one? Some will claim that it is determinism which is false. This position is called **libertarianism** (not to be confused with political libertarianism, which is a totally different idea). Others claim that determinism is true and that, therefore, there is no free will. This position is called hard determinism. A third type of position, **compatibilism**, rejects the incompatibilist definition of freedom and claims that free will and determinism are compatible (hence the name). The table below compares these different positions. But which one is correct? In the remainder of the chapter we will consider some arguments for and against these three positions on free will and determinism.

Is free will compatible with determinism?

		No	Yes
Is determinism true?	**No**	Libertarianism	
	Yes	Hard determinism	Compatibilism

Libertarianism

Both libertarianism and hard determinism accept the following proposition: If determinism is true, then there is no free will. What distinguishes libertarianism from hard determinism is the libertarian's claim that there is free will. But why should we think this? This question is especially pressing when we recognize that we assume a deterministic view in many other domains in life. When you have a toothache, we know that something must have caused that toothache and whatever cause that was, something else must have caused that cause. It would be a strange dentist who told you that your toothache didn't have a cause but just randomly occurred. When the weather doesn't go as the meteorologist predicts, we assume there must be a cause for why the weather did what it did. We might not ever know the cause in all its specific details, but assume there must *be* one. In cases like meteorology, when our scientific predictions are wrong, we don't always go back and try to figure out what the actual causes were—why our predictions were wrong. But in other cases we do. Consider the explosion of the Space Shuttle Challenger in 1986. Years later it was finally determined what led to that explosion ("O-ring" seals that were not designed for the colder condition of the launch). There's a detailed deterministic physical explanation that one could give of how the failure of those O-rings led to the explosion of the Challenger. In all of these cases, determinism is the fundamental assumption and it seems almost nothing could overturn it.

But the libertarian thinks that the domain of human action is different than every other domain. Humans are somehow able to rise above all of the influences on them and make decisions that are not themselves determined by anything that precedes them. The philosopher Roderick Chisholm accurately captured the libertarian position when he claimed that "we have a prerogative which some would attribute only to God: each of us, when we really act, is a prime mover unmoved. In doing what we do, we cause certain events to happen, and nothing and no one, except we ourselves, causes us to cause those events to happen" (Chisholm, 1964). But why should we think that we have such a godlike ability? We will consider two arguments the libertarian makes in support of her position: the argument from intuitions and the argument from moral responsibility.

The **argument from intuitions** is based on the very strong intuition that there are some things that we have control over and that nothing causes us to do those things except for our own willing them. The strongest case for this concerns very simple actions, such as moving one's finger. Suppose that I hold up my index finger and say that I am going to move it to the right or the to the left, but that I have not yet decided which way to move it. At the moment before I move my finger one way or the other, it truly seems to me that my future is open.

Nothing in my past and nothing in my present seems to be determining me to move my finger to the right or to the left. Rather, it seems to me that I have total control over what happens next. Whichever way I move my finger, it seems that I could have moved it the other way. So if, as a matter of fact, I move my finger to the right, it seems unquestionably true that I could have moved it to the left (and vice versa, *mutatis mutandis*). Thus, in cases of simple actions like moving my finger to the right or left, it seems that the strong incompatibilist

definition of freedom is met: we have a very strong intuition that no matter what I actually did, *I could have chosen otherwise, were I to be given that exact choice again*. The libertarian does not claim that all human actions are like this. Indeed, many of our actions (perhaps even ones that we think are free) are determined by prior causes. The libertarian's claim is just that at least some of our actions do meet the incompatibilist's definition of free and, thus, that determinism is not universally true.

The argument from moral responsibility is a good example of what philosophers call a **transcendental argument**. Transcendental arguments attempt to establish the truth of something by showing that that thing is necessary in order for something else, which we strongly believe to be true, to be true. So consider the idea that normally developed adult human beings are morally responsible for their actions. For example, if Bob embezzles money from his charity in order to help pay for a new sports car, we would rightly hold Bob accountable for this action. That is, we would punish Bob and would see punishment as appropriate. But a necessary condition of holding Bob responsible is that Bob's action was one that he chose, one that he was in control of, one that he could have chosen *not* to do. Philosophers call this principle **ought implies can**: if we say that someone *ought* (or ought not) do something, this implies that they *can* do it (that is, they are capable of doing it). The ought implies can principle is consistent with our legal practices. For example, in cases where we believe that a person was not capable of doing the right thing, we no longer hold them morally or criminally liable. A good example of this within our legal system is the insanity defense: if someone was determined to be incapable of appreciating the difference between right and wrong, we do not find them guilty of a crime. But notice what determinism would do to the ought implies can principle. If everything we ever do had to happen (think Laplace's demon), that means that Bob had to embezzle those funds and buy that sports car. The universe demanded it. That means he couldn't not have done those things. But if that is so, then, by the ought implies can principle, we cannot say that he ought not to have done those things. That is, we cannot hold Bob morally responsible for those things. But this seems absurd, the libertarian will say.

Surely Bob was responsible for those things and we are right to hold him responsible. But the only we way can reasonably do this is if we assume that his actions were chosen—that he could have chosen to do otherwise than he in fact chose. Thus, determinism is incompatible with the idea that human beings are morally responsible agents. The practice of holding each other to be morally responsible agents doesn't make sense unless humans have incompatibilist free will—unless they could have chosen to do otherwise than they in fact did. That is the libertarian's transcendental **argument from moral responsibility.**

Hard determinism

Hard determinism denies that there is free will. The hard determinist is a "tough-minded" individual who bravely accepts the implication of a scientific view of the world. Since we don't in general accept that there are causes that are not themselves the result of prior causes, we should apply this to human actions too. And

this means that humans, contrary to what they might believe (or wish to believe) about themselves, do not have free will. As noted above, hard determininsm follows from accepting the incompatibilist definition of free will as well as the claim that determinism is universally true. One of the strongest arguments in favor of hard determinism is based on the weakness of the libertarian position. In particular, the hard determinist argues that accepting the existence of free will leaves us with an inexplicable mystery: how can a physical system initiate causes that are not themselves caused?

If the libertarian is right, then when an action is free it is such that given exactly the same events leading up to one's action, one still could have acted otherwise than they did. But this seems to require that the action/choice was not determined by any prior event or set of events. Consider my decision to make a cheese omelet for breakfast this morning. The libertarian will say that my decision to make the cheese omelet was not free unless I could have chosen to do otherwise (given all the same initial conditions). But that means that nothing was determining my decision. But what kind of thing is a decision such that it causes my actions but is not itself caused by anything? We do not know of any other kind of thing like this in the universe. Rather, we think that any event or thing must have been caused by some (typically complex) set of conditions or events. *Things don't just pop into existence without being caused*. That is as fundamental a principle as any we can think of. Philosophers have for centuries upheld the principle that "nothing comes from nothing." They even have a fancy Latin phrase for it: *ex nihilo nihil fir* [1]. The problem is that my decision to make a cheese omelet seems to be just that: something that causes but is not itself caused. Indeed, as noted earlier, the libertarian Roderick Chisholm embraces this consequence of the libertarian position very clearly when he claimed that when we exercise our free will,

> "we have a prerogative which some would attribute only to God: each of us, when we really act, is a prime mover unmoved. In doing what we do, we cause certain events to happen, and nothing and no one, except we ourselves, causes us to cause those events to happen" (Chisholm, 1964).

How could something like this exist? At this point the libertarian might respond something like this:

> I am not claiming that something comes from nothing; I am just claiming that our decisions are not themselves determined by any prior thing. Rather, we ourselves, as agents, cause our decisions and nothing else causes us to cause those decisions (at least in cases where we have acted freely).

However, it seems that the libertarian in this case has simply pushed the mystery back one step: we cause our decisions, granted, but what causes us to make those decisions? The libertarian's answer here is that nothing causes us. But now we have the same problem again: the agent is responsible for causing the decision but

---

1. Actually, the phrase was originally a Latin phrase, not an English one because at the time in Medieval Europe philosophers wrote in Latin.

nothing causes the agent to make that decision. Thus we seem to have something coming from nothing. Let's call this argument **the argument from mysterious causes.** Here's the argument in standard form:

- The existence of free will implies that when an agent freely decides to do something, the agent's choice is not caused (determined) by anything.
- To say that something has no cause is to violate the *ex nihilo nihil fit* principle.
- But nothing can violate the *ex nihilo nihil fit* principle.
- Therefore, there is no free will (from 1-3)

The hard determinist will make a strong case for premise 3 in the above argument by invoking basic scientific principles such as the law of conservation of energy, which says that the amount of energy within a closed system stays that same. That is, energy cannot be created or destroyed. Consider a billiard ball. If it is to move then it must get the required energy to do so from someplace else (typically another billiard ball knocking into it, the cue stick hitting it or someone tilting the pool table). To allow that something could occur without any cause—in this case, the agent's decision—would be a violation of the conservation of energy principle, which is as basic a scientific principle as we know. When forced to choose between uphold such a basic scientific principle as this and believing in free will, the hard determinist opts for the former. The hard determinist will put the ball in the libertarian's court to explain how something could come from nothing.

I will close this section by indicating how these problems in philosophy often ramify into other areas of philosophy. In the first place, there is a fairly common way that libertarians respond to the charge that their view violates basic principles such as *ex nihilo nihil fit* and, more specifically, the physical law of conservation of energy. Libertarians could claim that the mind is not physical—a position known in the philosophy of mind as "substance dualism" (see philosophy of mind chapter in this textbook for more on substance dualism). If the mind isn't physical, then neither are our mental events, such our decisions.

Rather, all of these things are nonphysical entities. If decisions are nonphysical entities, there is at least no violation of the physical laws such as the law of conservation of energy.[2] Of course, if the libertarian were to take this route of defending her position, she would then need to defend this further assumption (no small task). In any case, my main point here is to see the way that responses to the problem of free well and determinism may connect with other issue within philosophy. In this case, the libertarian's defense of free will may turn out to depend on the defensibility of other assumptions they make about the nature of the mind. But the

---

2. On the other hand, if these nonphysical decisions are supposed to have physical effects in the world (such as causing our behaviors) then although there is no problem with the agent's decision itself being uncaused, there would still be a problem with how that decision can be translated into the physical world without violating the law of conservation of energy.

libertarian is not the only one who will need to ultimately connect her account of free will up with other issues in philosophy. Since hard determinists deny that free will exists, it seems that they will owe us some account of moral responsibility. If, moral responsibility requires that humans have free will (see previous section), then in denying free will we seem to also be denying that humans have moral responsibility. Thus, hard determinists will face the objection that in rejecting free will they also destroy moral responsibility.

But since it seems we *must* hold individuals morally to account for certain actions (such as the embezzler from the previous section), the hard determinist

needs some account of how it makes sense to do this given that human being don't have free will. My point here is not to broach the issue of how the hard determinist might answer this, but simply to show how hard determinist's position on the problem of free will and determinism must ultimately connect with other issues in philosophy, such issues in metaethics[3]. This is a common thing that happens in philosophy. We may try to consider an issue or problem in isolation, but sooner or later that problem will connect up with other issues in philosophy.

Compatibilism

The best argument for compatibilism builds on a consideration of the difficulties with the incompatibilist definition of free will (which both the libertarian and the hard determinist accept). As defined above, compatibilists agree with the hard determinists that determinism is true, but reject the incompatibilist definition of free will that hard determinists accept. This allows compatbilists to claim that free will is compatible with determinism. Both libertarians and hard compatibilists tend to feel that this is somehow cheating, but the compatibilist attempts to convince us arguing that the strong incompatibilist definition of freedom is problematic and that only the weaker compatibilist definition of freedom—free actions are voluntary actions—will work. We will consider two objections that the compatibilist raises for the incompatibilist definition of freedom: the **epistemic objection** and the **arbitrariness objection**. Then we will consider the compatibilist's own definition of free will and show how that definition fits better with some of our common sense intuitions about the nature of free actions.

The **epistemic objection** is that there is no way for us to ever know whether any one of our actions was free or not. Recall that the incompatibilist definition of freedom says that a decision is free if and only if I could have chosen otherwise than I in fact chose, given exactly all the same conditions. This means that if we were, so to speak, rewind the tape of time and be given that decision to make over again, we could have chosen differently.

---

3. One well-known and influential attempt to reconcile moral responsibility with determinism is P.F. Strawson's "Freedom and Resentment" (1962).

So suppose the question is whether my decision to make a cheese omelet for breakfast was free. To answer this question, we would have to know when I could have chosen differently. But how am I supposed to know *that*? It seems that I would have to answer a question about a strange **counterfactual**: if given that decision to make over again, would I choose the same way every time or not? How on earth am I supposed to know how to answer that question? I could say that it *seems to me* that I could make a different decision regarding making the cheese omelet (for example, I could have decided to eat cereal instead), but why should I think that that is the right answer? After all, how things seem regularly turn out to be not the case—especially in science. The problem is that I don't seem to have any good way of answering this counterfactual question of what I would choose if given the same decision to make over again. Thus the epistemic objection[4] is that since I have no way of knowing whether I would/wouldn't make the same decision again, I can never know whether any of my actions are free.

The **arbitrariness objection** is that it turns our free actions into arbitrary actions. And arbitrary actions are not free actions. To see why, consider that if the incompatibilist definition is true, then nothing determines our free choices, *not even our own desires*. For if our desires were determining our choices then if we were to rewind the tape of time and, so to speak, reset everything—*including our desires*—the same way, then given those same desires we would choose the same way every time. And that would mean our choice was not free, according to the incompatbilist. It is imperative to remember that incompatibilism says that if an action is not free if it is determined (*including if it is determined by our own desires*). But now the question is: if my desires are not causing my decision, what is? When I make a decision, where does that decision come from, if not from my desires and beliefs? Presumably it cannot come from nothing (*ex nihilo nihil fit*). The problem is that if the incompatibilist rejects that anything is causing my decisions, then how can my decisions be anything but arbitrary?

Presumably an arbitrary decision—a decision not driven by any reason at all—is not an exercise of my freedom. Freedom seems to require that we are exercising some kind of control over my actions and decisions. If my action or decision is arbitrary that means that no reason or explanation of the action/decision can be given. Here's the arbitrariness objection cast as a **reductio ad absurdum argument:**

1. A free choice is one that isn't determined by anything, including our desires. [incompatibilist definition of freedom]
2. If our own desires are not determining our choices, then those choices are arbitrary.
3. If a choice is arbitrary then it is not something over which we have control
4. If a choice isn't something over which we have control, then it isn't a free choice

---

4. The term "epistemic" just denotes something relating to knowledge. It comes from the Greek work episteme, which means knowledge or belief.

5. Therefore, a free choice is not a free choice (from 1-4)

A **reductio ad absurdum argument** is one that starts with a certain assumption and then derives a contradiction from that assumption, thereby showing that assumption must be false. In this case, the incompatibilist's definition of a free choice leads to the contradiction that something that is a free choice isn't a free choice.

What has gone wrong here? The compatibilist will claim that what has gone wrong is incompatibilist's idea that a free action must be one that isn't caused/determined by anything. The compatibilist claims that free actions can still be determined, as long as what is determining them is our own internal reasons, over which we have some control, rather than external things over which we have no control. Free choices, according to the compatibilist, are just *choices that are caused by our own best reasons*. The fact that, given the exact same choices, I couldn't have chosen otherwise doesn't undermine what freedom is (as the incompatibilist claims) but *defines* what it is. Consider an example. Suppose that my goal is to spend the shortest amount of time on my commute home from work so that I can be on time for a dinner date. Also suppose, for simplicity, that there are only three possible routes that I can take: route 1 is the shortest, whereas route 2 is longer but scenic and 3 is more direct but has potentially more traffic, especially during rush hour. I am leaving early from work today so that I can make my dinner date but before I leave, I check the traffic and learn that there has been a wreck on route 1. Thus, I must choose between routes 2 and 3. I reason that since I am leaving earlier, route 3 will be the quickest since there won't be much traffic while I'm on my early commute home. So I take route 3 and arrive home in a timely manner: mission accomplished. The compatibilist would say that this is a paradigm case of a free action (or a series of free actions). The decisions I made about how to get home were drive both by my desire to get home quickly and also by the information I was operating with. Assuming that that information was good and I reasoned well with it and was thereby able to accomplish my goal (that is, get home in a timely manner), then my action is free. My action is free not because my choices were undetermined, but rather because my choices were determined (caused) by my own best reasons—that is, by my desires and informed beliefs. The incompatibilist, in contrast, would say that an action is free only if I could have chosen otherwise, given all the same conditions again. But think of what that would mean in the case above! Why on earth would I choose routes 1 or 2 in the above scenario, given that my most pressing goal is to be able to get to my dinner date on time? Why would anyone knowingly choose to do something that thwarts their primary goals? It doesn't seem that, given the set of beliefs and desires that I actually had at the time, I could have chosen otherwise in that situation. Of course, if you change the information I had (my beliefs) or you change what I wanted to accomplish (my desires), then of course I could have acted otherwise than I did. If I didn't have anything pressing to do when I left work and wanted a scenic and leisurely drive home in my new convertible, then I probably would have taken route 2! But that isn't what the incompatibilist requires for free will. As we've seen, they require the much stronger condition that one's action be such that it could have been different even if they faced exactly the same condition over again. But in this scenario that would be an irrational thing to do.

Of course, if one's goal were to be irrational and to thwart one's own desires, I suppose they could do that. But that would still seem to be acting in accordance with one's desires.

Many times free will is treated as an all or nothing thing, either humans have it or they don't. This seems to be exactly how the libertarian and hard determinist see the matter. And that makes sense given that they are both incompatibilists and view free will and determinism like oil and water—they don't mix. But it is interesting to note that it is common for us to talk about decisions, action, or even whole lives (or periods of a life) *as being more or less free*. Consider the Goethe quotation at the beginning of this chapter: "none are more enslaved than those who falsely believe they are free." Here Goethe is conceiving of freedom as coming in degrees and claiming that those who think they are free but aren't as less free than those who aren't free and know it. But this way of speaking implies that free will and determinism are actually on a continuum rather than a black and white either or. The compatibilist can build this fact about our ordinary ways of speaking about freedom into an argument for their position. Call this the **argument from ordinary language**. The argument from ordinary language is that only compatibilism is able to accommodate our common way of speaking about freedom coming in degrees—that is, as actions or decisions being more or less free. The libertarian can't account for this since the libertarian sees freedom as an all or nothing matter: if you couldn't have done otherwise then your action was not truly free; if you could have done otherwise, then it was. In contrast, the compatibilist is able to explain the difference between more/less free action on the continuum. For the compatibilist, the freest actions are those in which one reasons well with the best information, thus acting for one's own best reasons, thus furthering one's interests. The least free actions are those in which one lacks information and reason poorly, thus not acting for one's own best reasons, thus not furthering one's interests. Since reasoning well, being informed, and being reflective are all things that come in degrees (since one can possess these traits to a greater or lesser extent) and since these attribute define what free will is for the compatibilist, it follows that free will comes in degrees. And that means that the compatibilist is able to make sense or a very common way that we talk about freedom (as coming in degrees) and thus make sense of ourselves, whereas the libertarian isn't.

There's one further advantage that compatibilists can claim over libertarians. Libertarians defend the claim that *there are at least some cases where one exercises one's free will* and that this entails that determinism is false. However, this leaves totally open the *extent* of human free will. Even if it were true that there are at least some cases where humans exercise free will, there might not be very many instances and/or those decisions in which we exercise free will might be fairly trivial (for example, moving one's finger to the left or right). But if it were to turn out that free will was relatively rare, then even if the libertarian were correct that there are at least some instances where we exercise free will, it would be cold comfort to those who believe in free will. Imagine: if there were only a handful of cases in your life where your decision was an exercise of your free will, then it doesn't seem like you have lived a life which was very free. In other words, in such a case, for all practical purposes, determinism would be true.

Thus, it seems like the question of how widespread free will is is an important one. However, the libertarian seems unable to answer it for reasons that we've already seen. Answering the question requires knowing whether or not one could have acted otherwise than one in fact did. But in order to know this, we'd have to know how to answer a strange counterfactual—whether I could have acted differently given all the same conditions. As noted earlier ("the epistemic objection"), this raises a tough epistemological question for the libertarian: how could he ever know how to answer this question? And so how could he ever know whether a particular action was free or not? In contrast, the compatibilist can easily answer the question of how widespread free will is: how "free" one's life is depends on the extent to which one's actions are driven by their own best reasons. And this, in turn, depends on factors such as how well-informed, reflective, and reasonable a person is. This might not always be easy to determine, but it seems more tractable than trying to figure out the truth conditions of the libertarian's counterfactual.

In short, it seems that compatibilism has significant advantages over both libertarianism and hard determinism. As compared to the libertarian, compatibilism gives a better answer to how free will can come in degrees as well as how widespread free will is. It also doesn't face the arbitrariness objection or the epistemic objection. As compared to the hard determinist, the compatibilist is able to give a more satisfying answer to the moral responsibility issue. Unlike the hard determinist, who sees all action as equally determined (and so not under our control), the compatibilist thinks there is an important distinction within the class of human actions: those that are under our control versus those that aren't. As we've seen above, the compatibilist doesn't see this distinction as black and white, but, rather, as existing on a continuum. However, a vague boundary is still a boundary. That is, for the compatibilist there are still paradigm cases in which a person has acted freely and thus should be held morally responsible for that action (for example, the person who embezzles money from a charity and then covers it up) and clear cases in which a person hasn't acted freely (for example, the person who was told to do something by their boss but didn't know that it was actually something illegal). The compatibilist's point is that this distinction between free and unfree actions matters, both morally and legally, and that we would be unwise to simply jettison this distinction, as the hard determinist does. We do need some distinction within the class of human actions between those for which we hold people responsible and those for which we don't. The compatibilist's claim is that they are able to do with while the hard determinist isn't. And they're able to do it without inheriting any of the problems of the libertarian position.

Study questions

1. True or false: Compatibilists and libertarians agree on what free will is (on the concept of free will).
2. True or false: Hard determinists and libertarians agree that an action is free only when I could have chosen otherwise than I in fact chose.
3. True or false: the libertarian gives a transcendental argument for why we must have free will.
4. True or false: both compatibilists and hard determinists believe that all human actions are determined.

5. True or false: compatibilists see free will as an all or nothing matter: either an action is free or it isn't; there's no middle ground.
6. True or false: compatibilists think that in the case of a truly free action, I could have chosen otherwise than I in fact did choose.
7. True or false: One objection to libertarianism is that on that view it is difficult to know when a particular action was free.
8. True or false: determinism is a fundamental assumption of the natural sciences (physics, chemistry, biology, and so on).
9. True or false: the best that support the libertarian's position are cases of very simple or arbitrary actions, such as choosing to move my finger to the left or to the right.
10. True or false: libertarians thinks that as long as my choices are caused by my desires, I have chosen freely.

For deeper thought

1. Consider the Shopenhauer quotation at the beginning of the chapter. Which of the three views do you think this supports and why?
2. Consider the movie *The Truman Show*. How would the libertarian and compatibilist disagree regarding whether or not Truman has free will?
3. Consider the Tolstoy quote at the beginning of the chapter. Which of the three views does this support and why?
4. Consider a child being raised by white supremacist parents who grows up to have white supremacist views and to act on those views. As a child, does this individual have free will? As an adult, do they have free will? Defend your answer with reference to one of the three views.
5. Consider the eccentric neuroscientist example (above). How might a compatibilist try to show that this isn't really an objection to her view? That is, how might the compatibilist show that this is not a case in which the individual's action is the result of a well-informed, reflective choice?

FREE WILL SUPPLEMENT: QUANTUM INDETERMINACY AND THE LIBET EXPERIMENTS

Matthew Van Cleave

In this supplement we will consider two issues concerning how modern science impacts (or doesn't impact) the traditional free will debate: one concerning whether quantum indeterminacy in physics creates room for free will and one concerning how certain experiments in brain science (the Libet experiments) purportedly undermine free will. In both cases we will see that there is no simple argument from science to either support the existence or denial of free will. Rather, the scientific results leave the problem of free will unsolved.

Quantum indeterminacy

Quantum indeterminacy (QI) refers to a range of different phenomena in quantum physics (the physics that studies the subatomic level) where the physical facts themselves seem to be indeterminate. One of the most famous examples of QI is the **Heisenberg Uncertainty Principle** which says that it is not possible to know both the position and the velocity of subatomic particles (such as electrons)[1]. Importantly, this uncertainty is understood by physicists not as a lack of our ability to measure correctly, but as built into nature itself. This contrasts with the way that physicists such as Isaac Newton understood the world. According to the Newtonian understanding of the world, if we knew all of the forces on a particular object, we could predict exactly what would happen to that object—where it would go, for example. Consider a billiards ball knocking into another billiards ball on a pool table, for instance. The velocity of that second billiards ball and the exact trajectory that it would take could be calculated if we knew all of the initial conditions—for example, the force of the first ball hitting it (determined by its mass and velocity) and coefficient of friction of the felt of the pool table. In short, **Newtonian physics** is deterministic: if we knew all of the initial conditions of some system, we could predict everything that would happen in that system. This is exactly the idea behind the famous "**Laplace's Demon**" thought experiment (which was referenced in "The Problem of Free Will and Determinism" chapter). In contrast, in **quantum physics** even if we knew all of the initial conditions, we

---

1. There are other kinds of indeterminacy in quantum physics, including indeterminacies about energy/time, as well as indeterminacies regarding the components of spin angular momentum. Thanks to Anthony Kuchera for clarification on this point (personal communication).

would still not be able to predict the exact position and velocity of an electron because the indeterminacy is built into the structure of nature. That means that the indeterminacy is not merely **epistemological** (due to our lack of knowledge), but rather **ontological** (build into the nature of the physical system itself). And that, in turn, means that at the subatomic (quantum) level, reality is non-deterministic—that is, even given all the forces on an object, where exactly that object will end up is indeterminate.

How does QI bear on the problem of free will and determinism? Here is how some people have tried to do that. The idea is that if the universe *isn't* deterministic (in the way that Newton thought that it was), then this leaves room for (libertarian) free will to exist. Recall that the libertarian requires that determinism be false in order for there to be free will. However, even if determinism is false, that is not sufficient to establish free will. Here's the reason why. What is happening at the quantum level isn't anything like what we are looking for in free will, rather it is something more like simple indeterminacy.

Indeterminacy is simply the denial of determinism. However, what we are looking for in free will is not simply indeterminism, but rather control over our actions. If I were to randomly (and without any control over the matter) simply start squawking like parrot in the middle of class, this wouldn't be free will, even if that action were not determined by all of the preceding events going on in my brain. Rather, such a behavior would exemplify a terrifying lack of control over my behavior. Thus, indeterminism (which is all that quantum indeterminacy gets you) is not the same thing as free will[2]. Free will requires something more—but what? Well, that isn't a question that quantum physics is going to be able to answer for us, nor should that be its goal!

The Libet experiments

Whereas quantum indeterminacy is supposed to support the libertarian view of free will, the Libet experiments are supposed to undermine that view. The Libet experiments take their cue from a crucial distinction: the difference between it *feeling like one has free will* and *actually having free will*. It is possible that although at the conscious level we strongly feel like we have free will, that feeling is an illusion. This is what Daniel Libet's experiments are often taken by determinists to show by hard determinists.[3] Determinist's don't deny that their actions feel free; they just claim that that feeling is an illusion. In 1677 Baruch Spinoza said: "Men are mistaken in thinking themselves free; their opinion is made up of consciousness of their own actions, and ignorance of the causes by which they are determined."[4]

---

2. You can find this line of argument in a number of places, including Robert Kane, A Contemporary Introduction to Free Will. Oxford University Press, 2005.

3. A good example of this strategy can be found in Daniel Wegner's The Illusion of Conscious Will. MIT Press, 2003.

4. This passage is found in Spinoza's Ethics, which can be found online here.

Here is how the Libet experiments work. There is a device called an electroencephalogram (EEG, for short) that measures electrical signals produced by the brain. As scientists have observed the brain's electrical activity, they have noticed a consistent pattern which they call the "readiness potential" (or RP, for short). The RP is a rise in the electrical activity in the brain that always proceeds an action by an average of 550 milliseconds (ms). Libet brought participants into the lab and hooked them up to an EEG. He then had them watch a little clock whose hand moved around and around the face and then instructed them to flex their wrist and to note when they become aware of their conscious intention to flex their wrist. What Libet found was that participants became aware of their conscious intention to flex their wrist about 200 ms before they actually flexed their wrist.[5] Thus the RP was reliably occurring about 300 ms before one reported any conscious intention to carry out an action. And that means that the RP itself cannot be identical to one's conscious intention to carry out the action (in this case, the flexing of the wrist). But it also means that an unconscious brain event (that is, the RP) is involved in causing the flexing of the wrist (since the RP always precedes the flexing of the wrist).

That is what Libet found in his experiments. But how is it supposed to undermine the existence of free will? The assumption of the libertarian view of free will is that our actions our under our control. But since we cannot control something that we are not consciously aware of (such as unconscious brain events), it follows that *if* what is causing our actions is an unconscious brain event, then we are not really in control of our actions. Here's that argument in standard form:

1. An action is free only if its cause is under one's conscious control
2. The RP is not something under one's conscious control.
3. The RP is causing the flexing of the wrist.
4. Therefore, the flexing of the wrist is not something under one's control [from 2-3]
5. Therefore, the flexing of the wrist is not a free action [from 1, 4]

Here is how Daniel Wegner interprets the results of the Libet experiments:

> The RP could thus signal the occurrence of unconscious mental events that produce both the experience of wanting to move and the occurrence of actual movement. This possibility alerts us to the intriguing realization that conscious wanting…is a mental event that is caused by prior events. It seems that conscious wanting is not the beginning of the process of making voluntary movement but rather is one of the events in a cascade that eventually yields such movement (Wegner 2003, p. 55).

---

5. Here is how the experiment was practically carried out. An EMG (which measures electrical signals produced by muscles flexing) measured when participants flexed their wrists. When they did so, this also triggered a recording of the EEG. The muscle burst measured by the EMG created the point in time against which the previous events (verbal report of the conscious intention to flex was well as the RP) were mapped.

This is an interesting challenge to free will, seemingly grounded in good science. Does this show that humans don't really have free will? The philosopher Al Mele has argued that it doesn't.[6] Consider first that if Libet's experiment is supposed to show in general that humans don't have free will, then the kinds of actions that participants undertake in his experiments would have to be representative of the kinds of actions human beings undertake in general. Are they? First of all, an action such as flexing one's wrist is an extremely simple action—about as simple as you can get. In contrast, consider a decision one would make in solving a math problem or deciding whether or not one wants to continue in one's current relationship with one's partner, or planning a vacation for spring break. These kinds of actions obviously involve more complex decision-making and are much more likely to be influenced by our conscious thoughts than simple actions like flexing one's wrist. Second, Libet's participants were *explicitly instructed to try to be as arbitrary as they could* in choosing when to flex their wrists. Many of the actions that we would want to claim as our own (as our freely chosen actions) would be precisely those that are *not* randomly chosen. Rather, they would be one's where our choices are informed by our reasons. So there are two good reasons for thinking that whatever it is that is occurring in the Libet experiments is not in general representative of all human actions. In other words, even if we grant that *simple, arbitrary actions* are not within our conscious control, it doesn't follow that *every* human action is not within our conscious control.

But there are deeper challenges to the idea that the Libet experiments show that humans do not have free will.[7] A common interpretation is that the RP *is itself* the decision to move one's wrist that is being made unconsciously (since it precedes one's conscious awareness). There's a certain kind of sense to this since the electrical activity in the brain is hitting a peak at that point (that's just what the RP is). However, it is important to understand that there is activity that is itself giving rise to the RP. Why not treat that prior activity as "the decision"? On the other hand, why not treat the activity after the peak as "the decision"? Further, why not think that the RP is just a precursor to the decision and the decision itself is actually being made around 200 ms before the action? The argument against free will seems to come from the idea that a conscious decision must not be caused by any other antecedents. However, although there are some who have held a view like this (for example,, Roderick Chisholm), it seems a really implausible assumption to saddle the defender of free will with. Instead, one could see the conscious decision as itself arising from earlier unconscious activity in the brain. How else could it possibly work? And just because the conscious decision is itself caused by previous brain events, that doesn't mean that the conscious decision is causally impotent.

Libet himself envisioned this idea. The idea is that even if the RP were the unconscious decision being made, this still left room for a "conscious veto." What he meant by this is that a decision could be overridden in the

---

6. A great, accessible introduction to both the Libet experiments and Mele's criticisms of it can be found in Mele's short book, Free: Why Science Hasn't Disproved Free Will. Oxford University Press, 2014.

7. The following two paragraphs are a rough summary of the criticism of Libet in Mele 2003.

last second by a decision to *not* flex one's wrist. So conscious decisions could still have some causal power even if we grant that the RP is the decision. Libet claimed to have found data to support the idea of a conscious veto by instructing participants to prepare to flex their wrists at a preset time (on the hand of a clock) but to then to veto the intention to flex their wrists. What Libet found was a similar kind of RP, but that flattened out about 200 ms before the preset time at which they were supposed to flex their wrists. Libet claimed that this flattening out of the RP was due to the conscious veto. However, one begins to wonder at this point whether the instructions to both intend to flex one's wrist but then veto that intention is sufficiently confusing for participants to be puzzled as to what they are really doing. As Al Mele has noted, "how can a normal agent simultaneously be settled on [flexing their wrists at a preset time] and not [flexing their wrists at a preset time]."[8]

Mele has also raised a simple but important criticism of Libet's methodology which casts further doubt on the idea that the RP is the decision. Libet only collected data in cases in which participants actually flexed their wrist. So although it is true that every time there is a RP, the action occurs 550 ms after, we do not know whether there is an RP even if the action *doesn't* occur (since when the action didn't occur, no EEG data was collected/kept). Compare the following two statements:

> Every time there is an RP, there is an action

> Every time there is an action, there is an RP

For us to be able to conclude that the RP was the decision, we would have to be able to say that both were true. However, the data collected only allows us to verify the first statement, not the second statement. If there were an RP even when the action did not occur, then clearly we could not say that the RP was the decision. This is the kind of possibility that would be true if there was a conscious veto and as we have seen in the previous paragraph there is at least some evidence that this is true.

However, a deeper issue is whether it is correct to see the RP as "the decision." One reason to think it isn't is what I've suggested above: if what we have an evolving chain of causes in the brain, some conscious, some unconscious, then on what basis should we isolate some subset of those causes and call it The Decision? Rather, it seems more accurate to think of the whole chain, starting from the rise of the RP to the ultimately engagement of the motor neurons as the decision. Perhaps it is mistaken to think of decisions as kinds of things that could be isolated by milliseconds. Instead, perhaps the decision is something that unfolds over time. In that case, the conscious aspects of those brain states are also part of the decision and hence part of the cause of the action. These considerations point us in the direction of further empirical and philosophical questions

---

8. Al Mele, *Effective Intentions: The Power of Conscious Will*. Oxford University Press, 2009, p. 52.

about the nature of mental states and of mental causation. This is not surprising given that philosophical questions centrally involve the analysis of concepts.

Study questions

1. True or false: If quantum indeterminacy is true, then not everything is strictly determined.
2. True or false: Quantum indeterminacy is sometimes taken to undermine the idea of libertarian free will.
3. True or false: Libertarianism would be vindicated as long as determinism were shown to be false.
4. True or false: Libet's experiments are sometimes taken to show that determinism is true and that human lack libertarian free will.
5. True or false: Libet found that the RP (that is, readiness potential—a brain event that reliably proceeds our actions) proceeded subject's conscious awareness of their intention to act.
6. True or false: The fact that RP is seen as causing our conscious intentions is thought to undermine the concept of libertarian free will.
7. True or false: One criticism of the idea that the Libet experiments show that humans lack free will is that the kinds of actions used in the Libet experiments are simple rather than complex actions.
8. True or false: Libet's experiments were set up in order to be able to test the following claim: "Every time there is an action, there is an RP."
9. True or false: Libet's experiment leaves room for what he calls a "conscious veto" and this leaves some room for free will.
10. True or false: There is nothing problematic about identifying our decision to act with RP.

FREE WILL SUPPLEMENT: LIBERTARIANISM AND DUALISM

Matthew Van Cleave

It sometimes happens in philosophy that a position one takes on one issue has implications for another issue.[1] A good example of this is the relationship between the "libertarian" position in the free will debate and a position called "dualism" in the mind/body problem debate (see the chapter on the mind body problem in this textbook). Recall that the problem of free will concerns how there can be any room for free will in a deterministic physical world. If the physical world is deterministic and if the mind is just the brain, which is itself a physical system, then it seems that there is no way to escape the conclusion that all of our actions, which are routed through our minds, are themselves deterministic consequences of what precede them. Thus, since the libertarian claims that there is free will and that therefore determinism is false, she needs some account of how our minds can escape the deterministic realm of the physical world.

Modern philosophers such as Rene Descartes and Immanuel Kant were well aware of this problem and this is in part what led them to subscribe to dualism.[2] **Dualism** holds that the mind is not a physical thing but an **immaterial** thing. This means that for the dualist the mind is not the same thing as the brain. Rather, the mind exists in a different realm from the realm of physics. If the mind is immaterial then this means it is not subject to the deterministic laws of the physical world and this, in turn, leaves open the possibility for an incompatibilist (libertarian) account of free will to be true. If the causes of our actions do not issue from a physical system like the brain, but instead issue from an immaterial mind, then it is possible for us to say that I could have acted differently, even if all the same conditions were in place. Dualism insulates the mind from the physical realm and thus makes possible an incompatibilist, libertarian account of free will.

Of course, just because dualism supports the libertarian account of free will doesn't mean that dualism is true. As a theory, dualism stands or falls on its own merit, based on how satisfactory its answer to the mind body

---

1. Thank you to Christopher Schneck for suggesting that I make the connection between libertarian accounts of free will and dualism explicit.
2. It is not the only reason they were attracted to dualism, however. Within a Judeo-Christian worldview, the idea that humans have an immaterial soul was also an influence on modern philosophers like Descartes and Kant. Indeed, until the second half of the 20th century, dualism was the most commonly held position on the mind body problem among philosophers.

problem is. The point I am making here is that it is an attraction of a theory that it can help us resolve puzzles in other philosophical areas. That a theory about x can also help us to better understand some other thing y should be a factor in our overall assessment of philosophical theories. This is no less true in philosophy than it is in science.

Other philosophers question whether dualism really provides the kind of account a libertarian owes us. It seems to these philosophers that invoking dualism in order to explain how human actions are insulated from determinism is just trading in one mystery for another. The most central objection to dualism is the **interaction problem**: how can an immaterial, nonphysical mind interact with (for example, cause) things in the physical world? We really have no idea how this can occur and this has been a problem for dualism ever since Elisabeth, princess of Bohemia raised it in correspondence with Descartes.

CHAPTER 6: PHILOSOPHY OF RELIGION

PHILOSOPHY OF RELIGION

Matthew Van Cleave

Does God exist?

The question of whether god[1] exists may seem like a fairly straightforward one: either there is a god or there isn't. However, how one answers this question depends in large part on how one defines "god" and the fact that people have very different concepts of what god is complicates the matter. Muslims believe that the term "god" refers to a single unitary being that is all-powerful, all- knowing, and perfectly good. Christians believe that "god" refers to a Trinity— three different "persons" that are also (somehow) one being. Jews believe that "god" refers to a unity, like Muslims do, but disagree with Muslims that Muhammad was a prophet (a representative) of god. And this is only to mention the so-called "Abrahamic" traditions that uphold somewhat similar conceptions of god! Other traditions like the Buddhist, Hindu, or ancient Mesoamerican traditions uphold conceptions of gods that differ even more fundamentally from the Abrahamic conceptions. Traditions like the Hindu and ancient Greek religious traditions are polytheistic, meaning that they believe that there are many different gods, not just one. Some religious traditions tend to think of god not so much as a personal being, but as an impersonal force.

Thus, different religious traditions take the term "god" to refer to very different things. One might think that these different views are not incompatible— perhaps all of the gods of these different religions equally exist, so all of the religious traditions can be equally correct. However, that doesn't work because each tradition typically claims that *only* its particular god exists (or that *their* god is the most powerful). So although there is a sense in which all of these different traditions believe in god, they do not all believe in the *same* god since they believe in different, incompatible gods. The Christian believes in the Christian (triune) god but disbelieves in all of the other gods (Muslim, Hindu, Buddhist, Greek, and so on). The Muslim believes in the Muslim god but not in any of the others. And so on for every religious tradition.

---

1. Throughout this chapter the uncapitalized term, "god," will be used to specify the generic concept of a higher being, irrespective of religion. In contrast, since specific deities are proper names, those will be capitalized (for example, Zeus, the Trinity, Yahew). Many religious have both generic terms for a deity and specific terms for specific deities. For example, the Hebrew word, Elohim, is a generic term for god, as is the Muslim term, Allah. (For this reason, Christians in Arabic-speaking countries refer to god as Allah, even though they are Christians, not Muslims.) In contrast, the Hebrew term Yaweh is a proper name that refers to a specific god. You can think of this distinction as analogous to the distinction between a car and Chevrolet. The term "car" is not capitalized but the proper name "Chevrolet" is.

There is a popular atheistic argument that proceeds from this fact that each religious tradition disbelieves in the god of every other religious tradition. It's called the "**one god further" argument**. Imagine a list of all the gods that humans have ever believed in. This list would include lots and lots of gods, including ancient Aztec gods, the gods of the Australian aborigines, the Greek gods, and the Christian god, to name just a few. Suppose you were to ask a Christian whether he believes in the gods on that list. He will say that he disbelieves in every single god on the list except one—the Christian god. So the Christian disbelieves in all of these other gods that other human beings have fervently believed in. Next you ask the Muslim the same question and he'll give you a similar answer: he will disbelieve in every god on the list except one—the Muslim god. And so on and so forth for every other religion. The only difference between each of these different religious traditions and the atheist is that the atheist goes "one god further" than each of these other religions. So atheism isn't that strange of a position at all. In fact, every religious tradition is itself atheistic with respect to every *other* religious tradition's god. In other words, those religious traditions already acknowledge not only the plausibility but also the truth of the position that the atheist holds with respect to *all* gods since those different religious traditions hold this position with respect to *almost all* gods. Perhaps the atheist has the simpler explanation here.

However, suppose that we were not interested in whether the Christian god or the Muslim god or the Aztec god exists, but whether any god at all exists. It could turn out, for example, that none of the gods on our list of gods humans have believed in exist. And yet it might still be the case that some other god exists. For example, there could be a being that created the universe but that no religious tradition has ever conceptualized or worshipped. One way of thinking about the arguments for the existence of god that follow in this chapter is that they are aimed at establishing a much more general notion of "god," one that is compatible with many different religious traditions, but doesn't entail any one of them. Although the arguments that follow were developed mainly within the Abrahamic religious traditions (Judaism, Christianity, Islam), they don't necessarily establish (if sound) the existence of an Abrahamic god, but rather something much more general.

In the remainder of this chapter we will consider two different arguments for the existence of this more general notion of god: the teleological argument and the cosmological argument. In addition, we will consider two challenges to the existence of god: the problem of evil and the problem of religious diversity. The point of this chapter is not to convince you that god exists or doesn't exist, but rather to consider some influential attempts to argue for one conclusion or the other. As always, these issues go much deeper than I can possibly introduce here. But the arguments presented here are a very good place to start.

The teleological argument

The teleological argument is an argument that moves from considerations about design in the natural world to the existence of a designer. (The word "*telos*" in Greek refers to the notion of a goal or purpose.) The basic idea behind the teleological argument is that if we admit that universe contains design (whether in particular

objects, such as the eye, or in the organization of the whole earth or universe) then we must give an explanation of where that design comes from and the best explanation of where that design comes from is from a non-human designer. Below we will discuss a specific version of the teleological argument called the **analogical teleological argument**. The crux of this argument is an analogy between artifacts (objects made by human beings) and natural objects like the eye or the heart. The idea is that if we admit that the design of artifacts requires someone who designed them, then we must admit that design in natural objects (like the human eye) also requires someone who designed them. However, unlike the artifact, the designer of natural objects cannot be a human being, thus it must be some other non-human, intelligent designer.

Suppose while digging in the ground one day you found an old pocket watch. A question would naturally arise about where the watch came from in a very different way from the question of where, say, a rock or a clod of dirt come from. In the case of the rock, it would be plausible to say that it had just always been there, whereas in the case of the watch it doesn't seem plausible to say that the watch had always been there. The difference between the rock and the watch is that the watch is an **artifact**—something that is made by humans rather than made by natural processes—whereas the rock isn't. The watch has many different parts that have been arranged precisely and intentionally such that a certain goal or purpose can be achieved, namely so that by the steady movement of the hands on the watch face we are able to keep track of time.

Artifacts look different than natural objects. Artifacts often have an apparent structure and function to them that non-artifacts lack. Imagine, for example, that you are walking through the woods in the fall and see a bunch of leaves arranged on the ground according to color. The yellowish leaves are on one side of the arrangement and the reddish leaves are on the other side and the color of the leaves in the arrangement gradually shade-off from the yellow the red. Were you to encounter this arrangement, what would you think? Probably that some human being had intentionally arranged the leaves by color. Why?

Because there is an order, a pattern, a design whose best explanation seems to be in terms of someone's goal or purpose rather than in terms of wind, rain, and other natural forces. A naturally occurring pile of leaves would not be arranged in this way. Rather the colors would tend to all be mixed together, with some leaves being right side up and other leaves being upside down. It would be extremely unlikely that nature (leaves falling randomly, wind, and so on) would produce an arrangement of leaves like that described above. The best explanation for the color-coded arrangement is that some person did it for some purpose (perhaps for aesthetic reasons). In contrast, we don't need any such explanation for the random pile of leaves since nature produces such effects all the time.

The pocket watch (imagine a mechanical device, not a digital one) has a similar kind of design that makes the inference to a human designer irresistible. The watch has gears that turn one another and that ultimately turn the second, minute, and hour hands on the watch face. It has a tiny wheel that allows you to wind it up. It has markings on the watch face that allows you to precisely tell where the hands are on the watch face. All of these features are analogous to the leaves arranged according to color in the above example. They are features that

exhibit design and that make the inference to a human designer feel irresistible and obvious in a way that the pebbles on the beach or the random pile of leaves isn't.[2]

One of the most famous examples of the analogical teleological argument comes from William Paley who was an Anglican minister and theologian at Cambridge University in the late 18th century. The watch example discussed above derives from Paley. Paley argues that if it is obvious that the watch was designed then it is just as obvious that certain natural objects were designed. One of my favorite examples of Paley's is one where he compares a telescope to an eye:

> As far as the examination of the [telescope] goes, there is precisely the same proof that the eye was made for vision, as there is that the telescope was made for assisting it. They are made upon the same principles, both being adjusted to the laws by which the transmission and refraction of rays of light are regulated. I speak not of the origin of the laws themselves; but such laws being fixed, the construction in both cases is adapted to them. For instance, these laws require, in order to produce the same effect, that the rays of light, in passing from water into the eye, should be refracted by a more convex surface, than when it passes out of air into the eye. Accordingly, we find that the eye of a fish, in that part of it called the crystalline lens, is much rounder than the eye of terrestrial animals. What plainer manifestation of design can there be than this difference? What could a mathematical instrument-maker have done, more, to show his knowledge of his principle, his application of that knowledge, his suiting of his means to his end; I will not say to display the compass, or excellence of his skill and art, for in these, all comparison is indecorous, but to testify counsel, choice, consideration, purpose? (Paley, *Natural Theology*, 1802)[3]

Paley is claiming the order and purpose observable in the telescope is analogous (the same in relevant respects) to the order and purpose is observable in natural objects, like eyes. If we find the inference to a designer irresistible in the case of the telescope, we should find it irresistible in the case of eyes (and a myriad of other natural objects that exhibit design), as well. The only difference is that whereas in the case of artifacts like watches, telescopes, and color-coded piles of leaves it is clear that humans could be the designers, in the case of natural objects like eyes, the designer could *not* be a human. Thus, in the case of natural objects, the designer would have to non-human. And what kind of thing might a non-human intelligent designer be (asks Paley with a twinkle in his eyes)?

As noted, Paley's argument is an **analogical argument** and the analogy explained in the following diagram.

---

2. Of course, it could have been the case that a person made the random pile of leaves and the assortment of pebbles on the beach, as well. However, although this could be the case, it also needn't be. In contrast, in the case of the watch and the color-coded pile of leaves, we really can't give a plausible explanation without invoking some human designer since those things don't just randomly assemble themselves without the intervention of a human being.

3. http://darwin-online.org.uk/content/frameset?itemID=A142&pageseq=1&viewtype=text
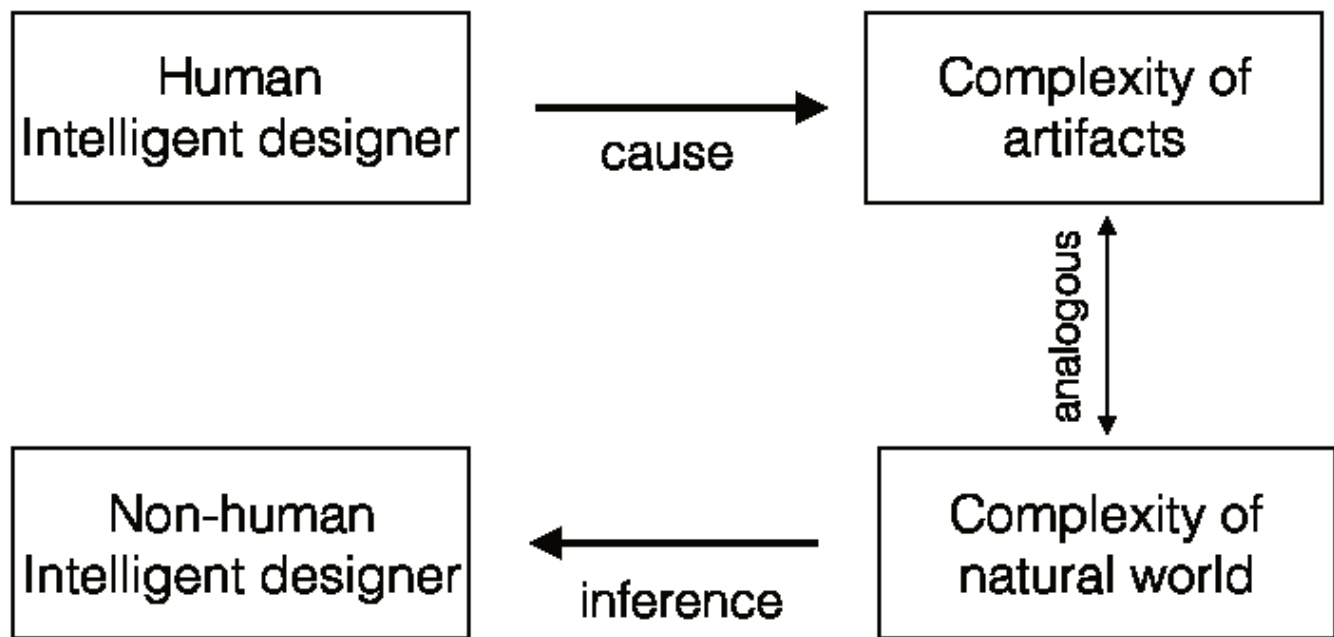
The complexity of artifacts is analogous to the complexity observed in natural objects. Thus, because we know that artifacts are designed by human intelligent designers, it must follow that objects in the natural world are also designed, albeit by some non-human natural designer. Here is Paley's reconstructed argument:

1. The examples of complexity we observe in artifacts (for example, a watch, a telescope, a color-coded pile of leaves, and so on) are created by human intelligent designers
2. We observe things of even more complexity in the natural world (for example, eyes, organisms, the universe itself)
3. Like effects imply like causes
4. Therefore, the cause of complexity in the natural world must a non-human intelligent designer of some sort (from 1-3)

Of course, Paley himself would like the ultimate conclusion to be that the god of Christianity exists. But I have suggested above that the best way of construing these arguments are for a much broader conclusion. The explicit conclusion of Paley's is just that there exists some non-human intelligent designer. One might hope that this gets you at least in the range of something godlike. However, it doesn't. Let's consider some different possibilities of what could count as "an intelligent designer of some sort."

Suppose that there is a civilization of hyper-intelligent Martians on some other planet in the universe. Suppose that this race of Martians created life on this planet as a kind of project (perhaps they've also done this on other planets in the universe that we don't know about). The Martians could be the non-human intelligent designer(s) that the argument's conclusion makes reference to. This seems a far cry from any notion of god. Rather, the Martians are just another race of intelligent beings in the universe that live and die and make

mistakes; they are not gods. Although one might try to define them as gods, the Martians seem to fall short of what most religions intend when they venerate and worship god.

In 1859 Charles Darwin published his magisterial *On the Origin of Species* in which we put forward and meticulously documented the theory of natural selection. What Darwin showed (and what has since been confirmed as well as any scientific theory has been) is that order and complexity can emerge from unintelligent, mechanical processes. All that is needed is several conditions,

including a mechanism of inheritance that is almost, but not quite, perfect. That is, the mechanism needs to reproduce organisms that have the traits of the parent organisms in a way that is almost, but not quite, flawless. We now understand exactly what this mechanism is in a way that Darwin didn't: it's DNA, which Watson and Crick discovered the structure of in the 1950s. It sometimes happens that DNA does not perfectly replicate and when this happens a mutation occurs. Sometimes these mutations can have beneficial effects for an organism and when they do, those mutations tend to be passed on to further generations and in this way the mutation tends to spread through a population. For example, suppose a moth was always light colored but then something changed in the environment to make the moth's environment darker colored. In this new environment, the light colored moth is more visible to predators and thus more likely to be eaten by them. However, suppose that a random mutation occurs in one of the moths to turn it from light-colored to dark- colored. In this new environment, the dark-colored moth will tend to be more successful than the light-colored moths (that is, it will tend not to get killed by predators as often as the light-colored moths). Because of how it enables the organisms that possess it to be successful, this new mutation that causes the dark coloring will tend to pass through a population.[4] In this way, species can change and, eventually, create new species. Darwin's idea is that by this same kind of process, over long periods of time, speciation occurs. Ultimately we get organisms that are amazingly well designed to their environments—exactly the kind of design that so struck Paley. However, although there is a sense in which organisms are indeed designed by natural selection, this design is not "forward looking" or intelligent. Rather, species have the traits they have simply because of random mutation combined with the process of natural selection. But this "designer" doesn't look at all like the non-human designer that Paley was intending. That is, natural selection doesn't look anything like a god.

Let's pause for a second to note an interesting connection between Paley and Darwin. Darwin actually had to read, and was much influenced by, Paley when he was a college student. Here's Darwin from his autobiography:

> Again in my last year I worked with some earnestness for my final degree of B.A., and brushed up my

---

4. This example is actually not made up, but real. You can read about it here: http://www.mothscount.org/text/63/peppered_moth_and_natural_selection.html

classics together with a little algebra and Euclid, which latter gave me much pleasure, as it did whilst at school. In order to pass the B.A. examination, it was, also, necessary to get up Paley's 'Evidences of Christianity,' and his 'Moral Philosophy.' ... The logic of this book and as I may add of his 'Natural Theology' gave me as much delight as did EuclidI did not at that time trouble myself about Paley's premises; and taking these on trust I was charmed and convinced by the long line of argumentation. ... The old argument of design in nature, as given by Paley, which formerly seemed to me so conclusive, fails, now that the law of natural selection has been discovered. We can no longer argue that, for instance, the beautiful hinge of a bivalve shell must have been made by an intelligent being, like the hinge of a door by man.[5]

Thus, the analogical teleological argument was delivered a huge blow by Darwin. The problem is that even if we agree that design implies a designer, it doesn't follow that that designer is intelligent or a person or any of the attributes that the religious traditions give to god. Essentially what Darwin did was that he challenged premise 3 of Paley's argument. His claim was that like effects do not always imply like causes and he showed that this was the case. When you think about it, premise 3 of the above argument seems to be the weakest. There are plenty of examples where two of the same types of things have very different causes. I'll leave it to the reader to come up with examples.

So not only does the analogical teleological argument fail to establish any particular conception of god (which in any case isn't best seen as its intention), it doesn't even establish a broad conception of god. On the one hand, the Martians example shows that even if we accept the validity of the argument, the conclusion doesn't establish that this non-human intelligent designer is godlike. On the other hand, natural selection shows that the whole analogy on which the argument is based is flawed since the existence of complexity and design does not establish that there was an intelligent designer. Natural selection is indeed a kind of designer, but it is not intelligent and does not seem to be in any kind of way godlike.

But this isn't even the last of the problems that have been famously raised against the analogical teleological argument. We will close this section with one last famous problem that was raised by David Hume before Paley ever devised his argument. What Hume noticed is that if complexity establishes that something is designed and if the *only* thing that can explain design is a designer, then an **infinite regress** looms for the teleological argument. What Hume recognized was that an intelligent designer is itself a complex thing and thus would *require* explanation in terms of some other intelligent designer. But then the same point would apply to this new intelligent designer: it is itself a complex thing and thus would require explanation in terms of some further intelligent designer. Now you should see the problem: the premises of the argument lead to a paradox

---

5. https://www.gutenberg.org/files/2010/2010-h/2010-h.htm

of an infinite number of different intelligent designers.[6] But presumably there can't be an infinite number of different intelligent designers and in any case this seems to be at odds with what most religious traditions believe.

There are other teleological arguments besides the analogical argument. One that has been much discussed in recent years is called the **fine-tuning argument**. The fine-tuning argument focuses on the extremely narrow range of physical constants that make possible life as we know it. The idea is that given all the conditions that must simultaneously be in place in order for there to be life, it is extremely unlikely that these would have just occurred. Rather, God must have "fine-tuned" them so that the universe could support life—specifically, human life.

But what is meant by the claim that this universe—the one that supports life like ours—is highly improbable? If we are comparing one possible option with another, then each option (of the billions of different options) has an *equally* low probability. So our particular universe isn't any more special than other possible universes in this sense. When we think of the probability of our particular universe, and saying that it is improbable, we note our particular case only because we have an interest in it. It is striking to us that we might not have existed because we can think of all the cases in which we wouldn't have existed if something very small had been different. But from the point of view of the universe our human interests don't matter. Rather, the universe cares just as much about the possibility in which humans exist as about all the other possibilities in which we don't. They're all equally probable from the perspective of the universe.

Furthermore, it doesn't make sense to say that this universe is "more complex" than other possible universes because, again, complexity is something relative to our interests. We single out the various physical constants that are important from *our* perspective—that is, those that support our human existence.

Consider an analogy. What is the likelihood that you exist? Your individual existence is extremely improbable, yes, but this is only relative to your interest in existing. Your existence is just a likely (or unlikely) as anyone else's; it isn't any more or less probable than any other possible person that might have existed in your place. Any of the myriad possible (but not actual) individuals that might exist today have the same kind of objective

---

6. In Hume's *Dialogues Concerning Natural Religion* (1779), the character Philo says, "How therefore shall we satisfy ourselves concerning the cause of that being whom you suppose the author of nature, or, according to your system of anthropomorphism, the ideal world, into which you trace the material? Have we not the same reason to trace that ideal world into another ideal world, or new intelligent principle? But if we stop, and go no farther; why go so far? Why not stop at the material world? How can we satisfy ourselves without going on in infinitum? And, after all, what satisfaction is there in that infinite progression?" The character Cleanthes responds to Philo: "You ask me, what is the cause of this cause? I know not; I care not; that concerns not me. I have found a Deity; and here I stop my enquiry. Let those go farther, who are wiser or more enterprising." https://www.gutenberg.org/files/4583/4583-h/4583-h.htm

probability of existing. Someone was going to exist, it just happened to be you. Yes, that was unlikely, but so was every other possibility. Nevertheless, one of the possibilities was going to take place, no matter what. We can dramatize the point by considering a coin- flipping tournament. How probable it is that I win 10 consecutive coin flips?

Not very likely, of course. But suppose we arrange an elimination tournament with 10 rounds (this will involve 1024 contestants). In that case, someone *has to* win and thus someone will win 10 consecutive coin flips. The winner will think they are special, but really someone had to win; it just happened to be them.

The improbability of their winning the tournament doesn't require any fine- tuning![7] The same point applies to the existence of humans in the universe. Yes, it is unlikely that we exist, but so is every other possibility. It just happened to be us but that doesn't make use special from the perspective of the universe. The actual universe containing human beings is just one of a myriad of different possible universes. Perhaps most of those other universes do not contain intelligent beings. If so, then although we are right to think that our existence is improbable, we are wrong to think that this requires a radically different kind of explanation—one involving fine-tuning. On the other hand, perhaps many of those other universes do contain life, including intelligent beings. If so, then our universe is not really that special at all (and thus our universe doesn't require any kind of fine-tuning). Thus, regardless out how likely life is in other possible universes, it doesn't seem to provide any reason to bring in any intelligent designers.

Study questions

1. True or false: the teleological argument rests on an analogy between artifacts and natural objects.
2. True or false: One of the premises of the teleological argument is that god is a loving and powerful being
3. True or false: if there is a nonhuman intelligent creator of the world, then it follows that god exists.
4. True or false: Darwin's theory of natural selection can be seen as a challenge to the idea that "like causes imply like effects."
5. True or false: The analogical teleological argument doesn't establish that the Christian god exists, but it does establish that some god or other exists.
6. True or false: There is more than one version of the teleological argument besides the analogical version.
7. True or false: One problem with the analogical teleological argument is that the intelligent designer would itself require an explanation in terms of another intelligent designer, and so on.

---

7. This example comes from Daniel Dennett's 1995 book, *Darwin's Dangerous Idea* (Simon and Shuster), pp. 53-56.

For deeper thought

1. Sometimes one and the same outcome/effect can have radically different causes. For example, the human being playing chess and making decisions based on their conscious awareness vs. a computer playing chess using a series of algorithms. From the outside the behavior of the two things looks the same, but the cause is totally different (conscious thought vs. mindless algorithm). *Come up with another example where one and the same thing have two radically different causes.*

2. One of the characters named Cleanthes in David Hume's *Dialogues Concerning Natural Religion* makes the following rebuttal to the infinite regress objection: "You ask me, what is the cause of this cause? I know not; I care not; that concerns not me. I have found a Deity; and here I stop my enquiry. Let those go farther, who are wiser or more enterprising." Essentially his response is that it doesn't matter that there are further designers of the intelligent designer and that this doesn't undermine the point that there is a god. Do you think this is a good response? Why or why not?

The cosmological argument

Whereas the teleological argument draws on specific, often scientific, details about natural objects on Earth, the cosmological argument proceeds from more abstract considerations about the nature of cause and effect in general and applies those considerations to questions about the origin of the universe. The cosmological argument is probably the oldest of the arguments for the existence of god since we can trace versions of it back as far as Aristotle (384- 322 BCE). Aristotle argued for the existence of an "unmoved mover" based on considerations about causation. If any thing that is in motion was caused by something else in motion, then if we were to trace those things in motion back in time and assuming the chain of causes/effects is not infinite, then it follows that there must have been some first thing that moved other things but that was not itself moved by any other thing. This is what Aristotle called the **unmoved mover.**

Like the teleological argument, there are a number of different versions of the cosmological argument. Islamic philosophers in the middle ages translated Aristotle, studied his arguments, and then developed different versions of the argument. The work of those Islamic philosophers influenced Jewish and Christian philosophers to come up with their own versions of the cosmological argument.[8] In this section we will consider just one version of the cosmological argument—one that comes from one of the most influential

8. See Peter Adamson, *Philosophy in the Islamic World: A Very Short Introduction*. Oxford University Press, 2015.

medieval philosophers: Thomas Aquinas (1225-1274). Here is Aquinas in his own words (translated in English from the original Latin, of course):

> The second way is from the nature of efficient cause. In the world of sensible things we find there is an order of efficient causes. There is no case known (neither is it, indeed, possible) in which a thing is found to be the efficient cause of itself; for so it would be prior to itself, which is impossible. Now in efficient causes it is not possible to go on to infinity, because in all efficient causes following in order, the first is the cause of the intermediate cause, and the intermediate is the cause of the ultimate cause, whether the intermediate cause be several or one only. Now to take away the cause is to take away the effect. Therefore, if there be no first cause among efficient causes, there will be no ultimate, nor any intermediate, cause. But if in efficient causes it is possible to go on to infinity, there will be no first efficient cause, neither will there be an ultimate effect, nor any intermediate efficient causes; all of which is plainly false. Therefore it is necessary to admit a first efficient cause, to which everyone gives the name God.[9]

Aquinas's argument bears a striking similarity to Aristotle's. Read carefully through Aquinas's argument above and see if you can figure out the logic of the argument: what is the conclusion? What are the premises and how are they supposed to support the conclusion? Obviously the main conclusion that Aquinas is driving at is that God exists, which is what Aquinas is concluding in the last sentence. However, it is important to note that the conclusion that the argument most directly establishes is that there is a "first efficient cause." He then goes onto assume that that first efficient cause is god. Aquinas doesn't really give much of an argument for that assumption (that the first efficient cause is god) and so it is obviously the weakest part of the argument.

But let's set that aside for now and look at the conclusion that he *does* provide an argument for. One of Aquinas's premises is just that a thing cannot be the efficient cause of itself (see the third sentence of Aquinas's argument above). All he means by "efficient cause" is just what we mean by "cause" (more or less).

Think of billiards balls knocking into each other: the first one bumps into the second and the second begins to move and then bumps into the third, and so on. That's the kind of thing that Aquinas is talking about here. And his point is that a thing cannot cause itself because it is not possible for a thing to be prior to itself. Consider the billiards balls again: the second ball cannot be the cause of its own movement because the second ball's movement required something else to be happening *before* it moved—in this case, the first ball's movement. That is, it would have required the second ball to be moving before it was moving, which is not possible. This seems to be generally true of causes and effects: if A is the cause and B is the effect, then A must

---

9. This passage is from Aquinas's *Summa Theologica*, which you can find online here: http://www.newadvent.org/summa/1002.htm

precede B in time. If A comes after B in time then A cannot be the cause of B. So, Aquinas's first crucial premise is: *it is not possible for something to cause itself*.

The next crucial premise that Aquinas attempts to establish in the above paragraph is that it is not possible for the chain of causes/effects to go on to infinity. What he means is just that any actual series of causes and effects will not be infinitely long. Rather, it will terminate at some point—at the beginning. Aquinas's reason for thinking that the chain of causes cannot be infinite is that if you were to remove the first cause then you'd have no other effects that follow. It might be helpful here to think of a line of dominoes: if you remove any of the dominoes from the chain, then the dominoes that follow will no longer be knocked over. Likewise, if you never knock over the first domino, then none of the others fall. Aquinas's point is this: we can see that things are being caused right now in a myriad of different ways. Those things must have causes and those causes must themselves have causes, and so on. But that couldn't be the case if there wasn't a first cause. Aquinas's crucial assumption seems to be that *if a chain of causes were literally infinite then that is the same as removing the first cause*. And as we have seen, if you remove the first cause (for example, you don't knock over the first domino) then none of the other causes will occur, which means that the effect that we are observing in the world right now won't occur either.

The only other premise that Aquinas is relying on in the above paragraph is the obvious one that there *are* things that are caused. This is obvious if we just observe the world around us. But Aquinas needs this premise because even if the above two crucial premises were true, it wouldn't mean that there had to be an uncaused first cause. Often times a premise will be unstated in an argument because it is already obviously true and doesn't need to be asserted. Other times the unstated premise is needed to make the argument work, but it is far from obviously true. I've denoted these unstated premises in the final reconstructed argument below as "missing premises." Here, then, is Aquinas's argument reconstructed with all the missing premises that are needed to make it valid:

1. There exist things that are caused. [missing premise]
2. It is not possible for something to cause itself.
3. The chain of causes of any event cannot be infinite.
4. Therefore, there is a cause of the existence of some things which is itself uncaused. [from 1-3]
5. If there is a cause of the existence of some things which is itself uncaused, then God exists. [missing premise]
6. Therefore, God exists. [from 4-5]

The logic of this argument is airtight, it seems. That is, if we assume the truth of the premises, the conclusion must be true. That is what it means for an argument to be valid. However, even if the argument is valid it might have a false premise. It seems there are two premises here to challenge: premise 3 and premise 5. Let's consider each premise in turn.

Premise 3 is essentially saying that the universe cannot be infinitely old. But is this true? Why couldn't the universe (and hence the chain of causes) be infinite? This is actually a question that physicists have pondered and investigated. The standard answer the physicists give is that the universe is not infinite but had a beginning—something they refer to as the **Big Bang**. If we calculate backwards based on an estimate about the size of the universe and the rate at which the universe is expanding, we can come up with a surprisingly precise number: 13.7 billion years. That is the age of the universe according to physicists. But does that settle the matter? Does this consensus within physics show that Aquinas was correct and that the universe is not infinitely old?

Here is one reason to think not. Consider what happened before the Big Bang. Physicists will often say that this is not a question that makes sense to ask from the perspective of physics since space-time came into existence with the Big Bang. However, that raises the obvious question of whether physics settles all of the questions it makes sense to ask about the origin of the universe? Even if physics can't comment on what took place before the Big Bang, we can still ask the question of whether there was something before it. In any case, logically it seems that there could have been a previous universe before this one. It is hard to see how one could rule out that possibility. But that is exactly what Aquinas is trying to do. Is his argument successful? Is it true that if there wasn't a first cause then there wouldn't be any present causes (which is clearly not the case)? Here is a reason to think not. Although it is mind-boggling to consider the possibility that the universe does not have a beginning, this possibility wouldn't mean that there couldn't be present causes. Aquinas seems to be assuming something like that one could never travel through an infinite series of causes in order to arrive at the current causes that we observe. But this seems to assume that we need to be able to go back all the way in time to trace this series of causes leading up to the current effect that we are observing. Why make that assumption? Why not just say that there was always a previous cause? We can go back as far as we'd like and we will find an uninterrupted chain of causes leading up to the present. Why does it matter that we can't go all the way back in time? In fact, if the universe really is infinite—if there was another universe before the Big Bang and so on for an infinite amount of "universes"—then the "chain" of causes could indeed be infinite. Saying that the chain of causes is infinite is not the same as removing the first cause in a known series of causes.

Rather, it is just making that series infinitely long. Because Aquinas thinks that an infinite chain is just like removing the first cause in a known sequence of causes, he thinks it is contradictory. But if we challenge that assumption, then Aquinas's third premise is false and the argument would fail.

The other obvious premise to challenge is premise 5. This is the missing premise that would be needed in order to get to the conclusion that god exists. As I said above, Aquinas doesn't really seem to argue for the main conclusion, but if he were to do so, he'd have to defend something like premise 5. That is, he'd have to show us why the existence of an uncaused cause is the same thing as god. But it seems that there are plenty of counterexamples to that. For example, suppose that the universe is actually finite and that there aren't multiple universes that have existed before this one. In that case, the Big Bang itself (or "quantum fluctuations" or

whatever physicists determine it to be) would be an uncaused cause. But the Big Bang doesn't seem to be the kind of thing that religions worship as god. It certainly isn't the kind of thing that Christianity considers to be god. So there's a huge leap in logic there that is totally unsupported by any evidence.

Let's pause for a moment to make an important point about the preceding two arguments (teleological argument, cosmological argument). What these arguments purport to do is to establish the existence of god based on neutral evidence. It seems that neither one of the arguments we've considered does that. For example, Aquinas's argument at best only gets us to an uncaused first cause—Aristotle's unmoved mover. Aquinas thinks it plausible to say that this thing is god, but we have seen that that inference doesn't follow. If the Big Bang (or quantum fluctuations) were the uncaused first cause then this isn't anything close to traditional religious concepts of god. But one might plausibly claim a much weaker conclusion instead. Suppose we grant that the cosmological argument establishes that there is an uncaused first cause (and the teleological argument that there is a non-human designer). If one is already religious and believes in a god then they might plausibly say that the best explanation for these two things—the thing that best answers to those descriptions (uncaused cause, non-human designer)—is god. But that explanation would already assume the existence of god as the thing doing the explaining (*explanandum*). The traditional point of these arguments is to *establish* the existence of god, not to assume it. Nevertheless, treating god as the best explanation of these purported facts would be a good move for the theist to take here. The question then becomes whether or not the theist's explanation (that is, god) is a better explanation than the scientific explanations (natural selection, quantum fluctuations/Big Bang). This opens up a rather deep and important question—the relationship between scientific and religious explanations—that we won't broach here.

Before leaving this section, consider the following question: Does logic force us to admit that there *must* have been a cause of the Big Bang? A theist might claim that it does. After all, it seems that something cannot come from nothing. The theist might say that even if the Big Bang is the cause of the universe, god is nevertheless the cause of the Big Bang, making god the ultimate cause. Is the theist's position the more rational one here? It might seem so. After all, it feels incomplete to just say that the universe sprang into existence for no reason and without any cause. God as an explanation seems to give satisfying resolution.

However, on second though, is it really a satisfying resolution? Isn't the theist just inheriting the same problem: that of admitting that there is something that has no cause? What caused god, after all? If the theist's answer is (as it seems it must be) "nothing," then it seems that the theist is in no better place rationally than the atheist is.

The cosmological argument raises some of the deepest questions humans can ask: where did everything come from? Why is there something rather than nothing? Has the universe always existed or did it come to exist and if so, how? When it comes to explaining the origin of the universe it seems that any position we take will involve embracing things that it is difficult to imagine could be true. Suppose we say that the universe is infinite and so there was no first cause. This could be true, but it is a hard thing for our human minds to understand.

Suppose, on the other hand, that there was a beginning. In that case it seems that there will have to be some uncaused cause. Whatever you call that thing, whether god or quantum fluctuations, the human mind balks at the very thought of an uncaused cause. Thus, it seems that everyone is, in a certain respect, in the same boat regarding explaining the universe: however you proceed, you're going to have to accept something that seems false or paradoxical.

Study questions

1. True or false: The cosmological argument attempts to establish the existence of god based on considerations about the nature of causes/effects.
2. True or false: The cosmological argument has roots in ancient Greek philosophers.
3. True or false: One of the premises of Aquinas' argument is that nothing can cause itself.
4. True or false: One assumption that Aquinas' argument relies upon is that the existence of an uncaused first cause would mean that god exists.
5. True or false: If it is scientifically true that the universe had a beginning (such as the Big Bang), then this undermines Aquinas' argument.
6. True or false: Aquinas' argument proves that if there is an uncaused cause, then god exists.
7. True or false: Regarding explaining the origin of the universe, the theist is in a stronger position than the atheist.
8. True or false: The original intention of Aquinas' cosmological argument was to provide neutral evidence that would establish to anyone that god exists.
9. True or false: The claim that god provides the best explanation for the cause of the universe is a *stronger* claim than the claim that considerations about the cause of the universe proves that god exists.

For deeper thought

1. What is one reason for thinking that it is not impossible for the chain of causes/effects to be infinite?
2. What is the difference between saying that god best explains the existence of the universe and saying that the existence of the universe proves that god exists?
3. What difficult claim must someone who is defending an infinite universe accept? What difficult claim must someone who is defending a finite universe (with a first cause) accept? Which claim do you think is less probable?
4. Suppose that someone claimed that there must be a cause of the Big Bang since the Big Bang cannot have just come from nothing (and nothing comes from nothing)? How might one respond to this claim logically? (Hint: Is there any inconsistency in invoking the "nothing comes from nothing" principle and also claiming that nothing caused god?)

The problem of evil

The foregoing sections have considered arguments that attempt to establish that there is a god, but there are also arguments that attempt to establish that god doesn't exist. One ingenious attempt to do this is the problem of evil.

Many religious traditions (including all the "Abrahamic" traditions—Judaism, Christianity, and Islam) conceive of god as a being who is both omnipotent (all- powerful) and omnibenevolent (all-good or all-loving). But given that it seems that really horrible, evil things exist in the world—a world that was created and designed by god—it is hard to see how god could be both loving and all- powerful at the same time. Rather, it seems that given the existence of evil in the world, either god isn't all-powerful or god isn't all-good. This is what is called the **logical problem of evil**. The "logical" just refers to the fact that the following three statements are logically inconsistent (that is, they cannot all simultaneously be true):

A. God is omnipotent
B. God is omnibenevolent
C. Evil exists

David Hume captures the essence of the logical problem of evil succinctly when in his *Dialogues Concerning Natural Religion* the character Philo says,

> "Is he willing to prevent evil, but not able? then is he impotent. Is he able, but not willing? then is he malevolent. Is he both able and willing? when then is evil?"xxxxx

If we assume that god is all-powerful, then god would be able to do anything, including eliminate any evil and suffering from the world. The fact that evil exists, would seem to entail that god is not all-loving or good. On the other hand, if we assume that god *is* all-loving then god would never want there to be things like needless suffering in the world. But given that there does seem to be needless suffering, that would mean that although god wants there to be no suffering, he actually can't control the fact that there is needless suffering. But if god wants to but can't eliminate needless suffering, then god is not all-powerful. Thus, the problem is that A, B and C (above) cannot all be true. If we assume A and C are true, then B is false. If we assume B and C are true, and A is false.

There's another option, however. Perhaps A and B are both true but C is false— that is, perhaps there isn't any evil in the world. This of course should raise the question of what we mean by "evil." For our purposes, let's define evil as the existence of needless suffering—suffering that doesn't seem to serve any purpose. Does it make sense to deny that evil exists in this sense? Let me give you two examples to consider. Consider genetic diseases

like "Harlequin's Ichthyosis" where a person grows too much skin at too fast a rate which ultimately kills the infant typically within a matter of a few months. That is just one rare disease, but there are many diseases that cause a lot of suffering for the people who have them. Here's another example of a very different kind of evil. In 1984 Josef Fritzl asked his 18-year old daughter to help him carry a door into the basement of their home and then sealed her shut inside of it and then filed a missing persons report, claiming his daughter had disappeared. He proceeded to keep her locked as a prisoner in the basement for 24 years and fathered 7 children with her, all the while living out his life with his wife upstairs and her knowing nothing of what was happening. These two very different cases represent two very different kinds of evil: natural evil and moral evil.

Whereas moral evil concerns suffering that results from the actions of human agents (like Josef Fritzl or Adolf Hitler), natural evil concerns suffering that is not the result of human actions—things like cancer, forest fires, tsunamis, and so on. This distinction will be important in considering responses to the problem of evil below.

Responses to the logical problem of evil

There are basically two responses to the logical problem of evil: 1) deny that evil exists, 2) claim that the inconsistency is only apparent. Pursuing the first path seems a hard pill to swallow. The parent whose child was born with a rare genetic disease that causes their child much suffering won't be swayed by the claim that that suffering isn't really a bad thing. The suffering all around us immense, even if we don't always know about it. From the suffering of the victims of all kinds of abuse, to the suffering of children who die each day from curable illnesses, to the suffering of individuals who have been hit by natural disasters—all of these kinds of suffering are bad and the world would be a better one if there weren't so much of this suffering. That is in fact why so many people spend their lives trying to alleviate this suffering. It is hard to deny that the world would be better if it didn't contain the suffering it does.

The other response is the much more common one and that is simply to deny that it is inconsistent for there to be a omnipotent, omnibenevolent god, on the one hand, and evil in the world, on the other. A common way of showing that statements are not really inconsistent is by adding a further statement that explicitly resolves the inconsistency. Here is an example of that using a case that has nothing to do with the problem of evil. Consider the following two statements:

- Bob is the tallest human being ever to have lived
- Sherman is taller than Bob

It looks like those two statements are inconsistent: if Bob is the tallest then Sherman cannot be taller than Bob. However, suppose we added the following statement:

- Sherman is not a human being

By adding that statement we explicitly resolve any apparent inconsistency. If Sherman is a tree, for example, then it can both be true that Bob is the tallest human being and also that Sherman is taller than Bob. Can we do something similar with the problem of evil? The former University of Notre Dame professor of philosophy, Alvin Plantinga, has suggested that we can and his idea is basically this:

A. God is omnipotent
B. God is omnibenevolent
C. Evil exists
D. *God created a world containing evil and had a good reason for doing so*

Plantinga's point is that *if* D is true, then there is no logical inconsistency between A-C. Of course, this raises the question of whether D is true and Plantinga and other philosophers have argued that it is. The most common way of defending D is by presenting a version of what has become known as the **free will defense**. The basic idea behind the free will defense is that there is some greater good that cannot have been achieved without god allowing for there to be some evil. That greater good is moral goodness. A world without moral goodness, where people willfully undertake moral actions and develop their moral character is a better world, all things considered, than a world in which there is no moral goodness. But one of the things that is logically required for

moral goodness to exist is the possibility of human freedom—that is, human must have the choice to make either good or bad decisions. Consider: if humans did not have the capacity to make morally wrong decisions, then could we really praise them for being morally good? The crucial claim of the free will defense is that without the possibility of being able to make morally wrong decisions, moral goodness cannot exist. That's because moral goodness cannot exist without genuine choice, including the option of being able to act/choose the morally wrong thing. So the idea is that god created human with free will so that they would be able to develop and exhibit moral goodness. However, given that humans have free will, they will sometimes make the wrong choice and thus bring evil (unnecessary suffering) into the world. But is a world with moral goodness really better than a world with no unnecessary suffering? The theist will try to argue that it is by having us consider what a world would be like without free will. God could have created a world containing creatures who were essentially robots who could only do the right thing—creatures who were programmed such that they were incapable of ever do anything that is morally wrong. This would be a world without evil, but it would seem to be lacking something important from a world in which the creatures could willfully choose to do good (when they had the option of doing bad). Thus, in order to bring about a better world containing moral goodness, god created beings with free will who could also willfully choose evil over good.

There are a number of things that one could respond to this argument, but let's just consider one important thing before moving on. The free will defense only applies to explaining why there are certain kinds of evil—what we have above called "moral evil." The free will defense does give us any reason for why there

would be "natural evil" since human free will is in no way involved in the suffering that results from things like Harlequin's Ichthyosis or a tsunami. In fact, there's a whole natural world of suffering that exists quite independently of the realm of human actions. Consider, for example, predation in the animal kingdom. There is immense suffering within the natural world and this is something that religious people have long been aware of since it sits oddly with the idea of a loving creator. Why would a loving creator god create a world predicated on suffering and death? The study of biology reveals that evolution depends on death in order for species to evolve, but even if that it so, why need there be so much suffering and predation along the way (and for that matter why need there be evolution at all)? Presumably the gazelle doesn't enjoy being eaten alive by the lions. The prophet Isaiah seemed to be aware of this when he envisioned a world in which there was no longer any predation—a more perfect world:

> The wolf shall dwell with the lamb,
> and the leopard shall lie down with the young goat,
> and the calf and the lion and the fattened calf together;
> and a little child shall lead them.
> The cow and the bear shall graze;
> their young shall lie down together;
> and the lion shall eat straw like the ox.[10]

The question is: why wouldn't a loving, all-powerful god have created such a world in the first place? The free will defense doesn't provide any answer to this question. It only gives a good reason for why god would allow moral evil, but does not provide any answer to why god allows natural evil.[11]

Study questions

1. True or false: If we reject the idea that god is omnipotent then there is no logical problem of evil.
2. True or false: The logical problem of evil is that there are three statements that it seems a theist must believe but that those statements are mutually inconsistent.

---

10. https://biblehub.com/esv/isaiah/11.htm
11. The reader may wonder about the following kind of common response that theists give: "God has reason for allowing natural evil but human cannot understand it." The theist is entitled to Philosophy of Religion 138 give us a response and that may make a lot of theological sense from within their worldview. However, it seems that as soon as one gives this response, they have essentially given up on the project of offering reasons for their religious beliefs. I'm not saying that that isn't a legitimate thing that the theist could do (indeed, the philosopher Søren Kierkegaard is famous for defending a kind of view like this—what philosophers call "fideism." Rather, I'm noting that this response marks the theist refusing to engage in offering reasons for her beliefs.

3. True or false: By "evil" what is meant is unnecessary suffering.
4. True or false: There is no important philosophical difference between natural evil and moral evil.
5. True or false: A good example of moral evil would be a lion eating a gazelle alive.
6. True or false: The free will defense is a good response to the problem of natural evil.
7. True or false: Josef Fritzl is a good example of moral evil.
8. True or false: Harlequin's Ichthyosis is a good example of moral evil.

For deeper thought

1. According to the free will defense, what is the greater good that justifies why god allows evil in the world? Do you think that this really is a greater good? Why or why not?
2. Can you think of a good reason why god would allow/create *natural* evil?
3. Consider a young fawn that is burned in a forest fire and lies suffering for days until it finally dies. Presumably such things have happened many times in the history of the world (even if no one knew about it). Does the existence of events such as this pose a challenge for the existence of an all-powerful, loving god? Why or why not?
4. Consider a theist who responds to the problem of evil as follows: "God created/allowed evil in the world so that human beings would be able to work alongside god—a cooperative venture—to make the world a better place." Do you think this is a good response to the problem of evil? Why or why not?

The problem of religious diversity

Whereas the problem of evil is a metaphysical challenge to the existence of god, the problem of religious diversity is an epistemic challenge to the existence of god. Roughly, **metaphysics** concerns what *exists* and what is *true* about the world whereas **epistemology** concerns our *reasons or justification* for believing something to be true. Sometimes there are situations where x could be true but that we could never know that it is true. For example, in physics there is a concept of things that are within our "light cone" which is basically the idea of things that we could physically access by travelling at the speed of light.

However, there are also things outside our light cone because those things are travelling away from us at faster than the speed of light. Hence it is not physically possible to access those things (not even with technology). Now consider the idea that there may be life other places within the universe but that is not within our light cone. We cannot rule out the existence of such life (that is, we cannot rule it out metaphysically) but we could never know that there is such life. Thus the possibility that there is life outside our light cone presents an epistemic problem, not a metaphysical one. The problem of evil is a *metaphysical challenge* because it claims

that it is not possible for an omnipotent, all-loving god to exists (given evil). In contrast, as we will see, the problem of religious diversity presents an *epistemic challenge* to believing in a specific version of god within a specific religious tradition.

The problem of religious diversity asks us to consider religious believers from different traditions who believe in different gods and have different associated religious beliefs. For example, Christians believe that god is a Trinity (three persons but also, mysteriously, only one being) whereas Muslims believe that god is a unity and find the Trinity belief heretical. Judaism, in contrast, does not take Jesus to be divine, although they do take him to have been an important rabbi. Islam takes Jesus to be an important prophet but, unlike Christians, do not believe that Jesus was god or that his message was god's most important prophet. Instead, Islam takes Muhammad to be god's most important prophet. If we were to bring in non-Abrahamic religions, the conception of god would begin to diverge even more radically. All of these different religious traditions believe that their tradition is correct and this entails that the others are mistaken somehow. It is this fact that creates the epistemic problem, however. For consider a religious believer from a different tradition. For example, if you are a Christian, consider a Muslim who is just as intelligent, just as well-informed, just as devout as you are. Further, the Muslim has her own religious experiences,

draws on her own religious traditions, including her own religious texts and authorities. The Muslim, of course, believes that her tradition is the correct one just as much as you believe that your religious tradition is the correct one. Even if you are polite to each other about such things (as you should be), you probably still hold this belief and don't explicitly express it to each other.

Nonetheless, the belief is there. But consider the kind of position these religious believers are in by claiming that the other is fundamentally mistaken about some important religious truth. You would have to be claiming that someone who has is in a very similar epistemic situation to you is actually fundamentally mistaken! Should this not raise the question of whether *you* are fundamentally mistaken if they are? Consider the tension of the Christian's position:

I accept that my Muslim friend is just as intelligent and well-informed as I am and that they have their own religious tradition and religious experiences that they have come to trust just like I have and yet my beliefs are the correct ones and theirs are incorrect.

It seems that this recognition should occasion some doubt about the veracity of our own beliefs. In short, if the Christian accepts that a Muslim could have all the same kinds of evidence that they themselves do (religious experiences, religious traditions, religious texts, and so on) and yet that the Muslim is profoundly mistaken in their religious beliefs, then this admission should raise a question about the status of the Christian's *own* beliefs. If my friend can have all the same kinds of evidence that I do and yet be profoundly mistaken, then perhaps I am mistaken too! Thus, the crux of the problem of religious diversity is that in admitting there are well-informed, reasonable, and devout believers within different religious traditions we thereby undercut our

own epistemic position. That is, we acknowledge that possibility that we, too, are profoundly mistaken. And that recognition should make us much less certain that our beliefs are correct.

Responses to the problem of religious diversity

There are generally three different responses to the problem of religious diversity: religious pluralism, agnosticism/atheism, and religious exclusivism. **Religious pluralism** involves jettisoning specific religious beliefs attached to specific religious traditions and upholding a more general view of god that isn't specific to any religious tradition (or at least that incorporates all the consistent parts of the different religions). For example, the Christian, were she to become a pluralist, would have to renounce the beliefs that Jesus was literally god, that there is a heaven an hell, that god is a Trinity, that the Bible (specifically the New Testament) is the unique revelation of god to the world, and so on. A well- known proponent of this kind of view was John Hick.[12] The pluralist hopes that by moving to a more generalized conception of a transcendent being, she can avoid the problems that attach to making the more specific claims of the different religious traditions. It is often objected to this view that it collapses into a kind of agnosticism, according to which we can't really know anything at all about the supposed deity. **Agnosticism** and **atheism** involve jettisoning any belief in a god, finding it more plausible that we are all fundamentally mistaken about believing that there is any kind of god or supernatural being. Finally, religious exclusivism involves doubling down one one's specific religious beliefs by claiming that there is some good reason for the religious believer to hold on to their specific beliefs. It is important to recognize two very different versions of **religious exclusivism.** On the one hand, one might claim that there are **pragmatic reasons** to hold onto one's religious tradition. For example, if all of your family and friends are Christians and all of your social life revolves around that community and if becoming an atheist would alienate you from that community, then it would make sense to just remain a Christian rather than become an atheist. However, one should see that this kind of pragmatic reasoning doesn't bear at all on what we might call one's **epistemic reasons.**

Epistemic reasons concern evidence that something is true and it is clear that pragmatic reasons only involve what makes it easier for us to attain our goals. (Here's a quick example to illustrate the difference. Consider a mother whose husband and father of her young children has died in some ignoble way, say as a result of erotic asphyxiation. The mother has a very good pragmatic reason to not tell her children the truth about the death of their father because of the way this would adversely affect her children's wellbeing. But this pragmatic reason has nothing to do with what the evidence or truth actually is. Given her goal to protect her children, she lies to them. She has a good pragmatic reason to say, for example, that their father died by falling down the stairs but no epistemic reason for believing this.)

---

12. https://www.iep.utm.edu/hick/

Another version of religious exclusivism says that there are epistemic reasons for a religious believer to maintain their specific religious beliefs/tradition. Well- known proponents of this view in recent years include William Alston and Alvin Plantinga (mentioned earlier). However, it is beyond the scope of this introductory chapter to engage these fascinating attempts to safeguard the epistemology of traditional religious beliefs (in this case, Christian beliefs).

Study questions

1. True or false: Metaphysics concerns what we have reason or justification for believing to be true.
2. True or false: Epistemology concerns what exists or what is true about the world.
3. True or false: The problem of religious diversity is an epistemic challenge to religious beliefs.
4. True or false: The problem of religious diversity assumes that there are intelligent, devout, well-informed persons within religious faiths other than our own.
5. True or false: One objection to religious pluralism is that it collapses into agnosticism.
6. True or false: pragmatic reasons and epistemic reasons are the same thing according to the religious exclusivist.
7. True or false: The religious exclusivist claims that only one religion is correct.

For deeper thought

1. Would someone who believed in a very general, non-specific "higher being" face the problem of religious diversity? Why or why not?
2. Does religious pluralism really avoid the problem of religious diversity? Suppose she finds a devout Christian who is equally intelligent and equally as well-informed as she is. Should this undermine her confidence in her religious pluralism? Why or why not?
3. Suppose that believing there is no god makes Mark very unhappy. His priest tells him that this gives him a good reason to try believing in god. Is the priest correct that this gives Mark a good reason to believe in god? Explain.
4. One argument for atheist is that it is a simpler explanation of the fact that different religions disagree about the nature of god. Whereas the religions maintain that there is a god, but disagree about the nature of god, atheists maintain that the simplest explanation of the disagreement is that there is no god and humans are simply projecting these beliefs onto the world (Freud's explanation of religion is similar to this). What do you think about this explanation of religious disagreement?

CHAPTER 7: THE ETHICS OF BELIEF

PRAGMATIC ARGUMENTS AND THE ETHICS OF BELIEF

Matthew Van Cleave

Pascal's wager

Suppose that there are no conclusive arguments either for the existence of god or against the existence of god. Suppose further that the evidence regarding god's existence is 50/50: there are as many reasons for disbelieving there is a god as there are for believing there is a god. Would it be rational for us to believe in god, even in the absence of compelling evidence that god exists? Or does the absence of compelling evidence mean that we should be agnostic?[1] This is the background to a famous argument from the philosopher and mathematician, Blaise Pascal (1623-1662). Pascal believed that human reason will not be able to settle the matter of whether god exists, but he also thought that whether one believes that god exists is a vital issue that concerns one's wellbeing—even if not in this life, then in the next. Pascal's claim is that it is in our rational self-interest to believe that god exists. Pascal's argument is often referred to as Pascal's Wager.[2]

Consider a commonly held Christian (and Muslim) view, according to which there is a heaven, where believers will go and experience infinite bliss after they die, and a hell, where non-believers will live an eternity in torment and suffering. Given that we do not know that this is true, Pascal's point is that we can still ask whether it is rational to *believe* that it is true. And Pascal's answer is that *it is rational to believe it because it is in our self-interest to believe it*. Here's the reasoning behind the wager: If you believe that god exists and god really does exist, then (after you die) you have an *infinite* gain. If you believe in god and god doesn't exist, then you may lose out on some earthly pleasures (as a result of living a good Christian life), but whatever losses those are will be *finite*. In contrast, if you do not believe that god exists but god really does exist, then you lose out on an infinite reward (and possibly incur an infinite loss). The only other option is that you fail to believe that god exists and god really doesn't exist. In that case, you may gain some earthly pleasures but those will still be

---

1. "Agnostic" comes from the Greek word "gnosis" meaning "knowledge." The prefix "a" in Greek means "no" or "not." Hence, "agnostic" means literally "no knowledge." An agnostic is a person who neither believes nor disbelieves something. Thus an agnostic concerning god is someone who neither believes nor disbelieves in god.
2. The wager occurs in Pascal's *Pensées,* which is basically a collection Pascal's journal entries— his "thoughts" ("pensées" is Latin for "thoughts").

finite in contrast with the potentially *infinite* pleasures of heaven. The figure below represents all the possible options.

	You believe in god	You don't believe in god
God doesn't exist	finite loss	finite gain
God exists	infinite gain	infinite loss

Pascal thinks that it is obvious that the rational thing to choose is to believe in god because of the potentially infinite reward, which would offset any possible finite losses or gains. Therefore, if you don't already believe in god, it is in your rational self-interest to believe.

Pascal's is giving us a **pragmatic argument:** he is not trying to convince us that god actually does exist, but that it is in our self-interest to *believe* that god exists. Even if god doesn't exist, it would still be pragmatically rational for us to believe that god exists, given the potentially infinite reward as compared to the finite gains/losses of not believing. **Pragmatic reasons** for believing something contrast with **epistemic reasons** for believing because the latter are tied to *truth* whereas the former are only tied to our *self-interest* or *goals*. One of the questions raised by Pascal's argument is whether we should (or even are able to) believe something for

pragmatic reasons. As we will see below, the American philosopher and psychologist, William James, agrees with Pascal that we can and should believe for pragmatic reasons, whereas the English mathematician and philosopher W.K. Clifford thinks that it is morally wrong to believe something for purely pragmatic reasons. We will consider their arguments below.

Pascal's wager is one of the earliest known examples of the mathematical field of decision theory. Decision theory attempts to describe, in precise mathematical terms, what it is rational to do. Here is a simple example to illustrate the idea.

> Example: lets suppose we're guessing what color of marble has been randomly selected (as in a lottery) from an urn of colored marbles. There are 50 red marbles, 10 green marbles and 5 blue marbles. Suppose we get $10 if a red marble is selected, $100 for a green marble and $500 for a blue marble.

According to decision theory, the rational option to choose is the one that has the highest expected utility. Expected utility of a choice is calculated by taking the choice's probability multiplied by the choice's payoff. Thus the expected utility of the different options in this example would be:

Red = (.5)(10) = 5
Green = (.1)(100) = 10
Blue = (.05)(1000) = 50

In this case betting that a blue marble has been selected has the highest expected utility and thus, according to decision theory, would be the most rational bet.

We can apply decision theory to more complicated, real-life examples, as well. All we would need is an agent's preference ranking of her possible choices and an estimation of the probability that each of those choices has of occurring. The rational thing to do will be the thing that has the highest expected utility.

Importantly, the rational thing to do will be relative to our different desires (preference rankings). I may assign a really high utility to A and a low utility to B, whereas you may do the opposite. In that case, what it is rational for me to do is not what it is rational for you to do.

What is unique to Pascal's wager is that some of the choices are supposed to have an *infinite* payoff (in this case, the eternal bliss of heaven, if it exists).

Although Pascal himself set the probability of god's existence at 50%, since the payoff is infinite, it doesn't matter how low we set the probability of god's existence. If the payoff is infinite then the expected utility is infinite even if the probability is low. For example, even if we set the probability of god's existence at one in one hundred billion, the infinite payoff would still yield an infinite expected utility: $(.00000000001)(\infty) = \infty$. Thus, although Pascal doesn't explicitly do so, one could try to extend his argument even to situations where god's

existence is much less likely than god's nonexistence.Note, also that even if we add in the costs of believing in god, as long as the costs are finite then the expected utility will still be infinite. We can set the costs of wagering against God at zero. Pascal thinks that there is little or no payoff for one who wagers against God.

There's a problem with the idea of believing for pragmatic reasons that Pascal anticipates and responds to and that is this: *one can't make oneself believe something that ones has no compelling reason to think is true*. For example, I cannot simply make myself believe that there is $100 in my pocket, even if I would very much like for it to be true that there is $100 in my pocket. Pascal's agrees that you cannot simply make yourself believe something you don't feel is true, but he thinks that you can do things that will change the way you feel. Here is what Pascal says:

> Endeavour then to convince yourself, not by increase of proofs of God, but by the abatement of your passions. You would like to attain faith, and do not know the way; you would like to cure yourself of unbelief, and ask the remedy for it. Learn of those who have been bound like you, and who now stake all their possessions. These are people who know the way which you would follow, and who are cured of an ill of which you would be cured. Follow the way by which they began; by acting as if they believed, taking the holy water, having masses said, and so on. Even this will naturally make you believe, and deaden your acuteness.

In short, Pascal's response is: *fake it until you make it*.

The ethics of belief

The British philosopher W.K. Clifford couldn't disagree more with this. According to Clifford, our beliefs have consequences and being careless with our beliefs can have horrible consequences on other people. Because this is true, Clifford argues that *we have an ethical duty to uphold stringent standards for what we allow ourselves to believe*. In short, Clifford thinks that we have a moral duty to proportion our belief to the evidence. Thus, if the evidence for a particular claim is not sufficiently strong, then we should not believe that claim. So in a case like Pascal's where the evidence for god's existence is equivocal, we should neither believe that god exists nor believe that god doesn't exist. That is, we should be agnostic with respect to god's existence. According to Clifford, only evidence should drive what we believe, never our self-interest or what we wish were true. What Clifford is rejecting, then, is what we have called "pragmatic arguments": arguments that attempt to establish that we ought to believe something even in the absence of sufficient evidence that that thing is true. Clifford's view has become known as the "ethics of belief" (after the title of Clifford's 1877 article). The **ethics of belief,** then, is the claim that we have an ethical duty to believe only those things for which we have sufficient evidence. The ethics of belief stands in diametric opposition to pragmatic arguments like Pascal's.

Clifford's argument for the ethics of belief turns on a story that he tells about a ship owner. Here is that story:

> A ship owner was about to send to sea an emigrant ship. He knew that the ship was old and not well built, that it had seen many voyages, and that it often had needed repairs. People had suggested to him that possibly the ship was not seaworthy. These doubts preyed upon his mind, and made him unhappy; he thought that perhaps he ought to have the ship thoroughly overhauled and repaired, even though this would cost him a lot of money. Before the ship sailed, however, he succeeded in overcoming these doubts. He said to himself that the ship had gone safely through so many voyages and weathered so many storms that it would probably come safely home from this trip also. He would put his trust in God, which of course would protect all these unhappy families that were leaving their fatherland to seek for better times elsewhere. He would dismiss from his mind all doubts about the honesty of the ship builders. In such ways he acquired a sincere belief that his ship was thoroughly safe and seaworthy; he watched her departure with a light heart, and warm wishes for the success of the exiles in their strange new home; and he got his insurance-money when she went down in mid- ocean and told no tales.[3]

What Clifford's ship owner is doing here seems to be exactly what Pascal's questioner (who doesn't believe in god) is doing: they are both stifling doubts and a lack of belief by undertaking actions that will enable them to believe. In Pascal's case, that is by acting as if one believed (going to mass, taking holy water); in the case of Clifford's ship owner, that is by a kind of hopeful self-talk that enables him to overcome evidence to the contrary (the ship is old, not well- built, people suggesting that it wasn't seaworthy). However, in the case of Clifford's ship owner the action of stifling doubts is at the cost of people's lives and thus it seems that what the ship owner did was negligent and wrong. It doesn't matter that the ship owner believed that the ship would be fine; he had a duty not to believe it because of the evidence to the contrary. Here is what Clifford says:

> What shall we say of him? Surely this, that he was truly guilty of the death of those men. It is admitted that he did sincerely believe in the soundness of his ship; but the sincerity of his conviction does not exonerate him, because he had no right to believe on the evidence he had. He had acquired his belief not by honestly earning it in patient investigation, but by stifling his doubts. And although in the end he may have felt so sure about it that he could not think otherwise, yet inasmuch as he had knowingly and willingly worked himself into that frame of mind, he must be held responsible for it.

So the ship owner example is a clear case where a non-evidence-based belief (wishful thinking) led to a horrible outcome. If the ship owner had done his ethical duty he would have investigated whether the ship really was seaworthy rather than simply relying on wishful thinking. Thus it is wishful thinking that seems to be at fault here. As Clifford later notes, even if the ship had made a successful voyage, it still seems that the ship owner did something morally wrong because of the risk he put others in. Consider an analogy: even if a very drunk driver

---

3. Both here and in the following passages from Clifford, I have slightly updated Clifford's English to make it more modern and readable for students.

were able to make it home safely and didn't cause any harm, that doesn't make his action morally permissible. The fact is, he was putting others at risk and that is what makes that action wrong.

But what about cases where my wishful thinking is not harming others or even posing any substantial risk of harm? Return to Pascal's case of the person who doesn't believe that god exists but who is able to work themselves into that belief over time. What does it matter if this person believes in god or not? How is that hurting anyone? Or, to take a different case, consider someone who really, really wants unicorns to be real but doesn't currently believe they are real. In order to acquire this belief, this person spends most of their time with other people who also believe in unicorns and dismisses or ignores or explains away evidence to the contrary and basically does all the kinds of things the Clifford's ship owner does. Eventually, like Pascal's fake-it-until-you-make-it believer, this person comes to sincerely believe that unicorns exist. However, in this case it seems like a stretch to say that what this person has done is morally wrong because this person's belief doesn't seem to be harming anyone. What would Clifford say about a case of wishful thinking like this one? Isn't it importantly different from his ship owner case?

Clifford's response takes a step back to consider the general effects on society of believing things based on insufficient evidence. Clifford's claim is that by allowing ourselves to believe things without sufficient evidence, we weaken our powers of critical awareness upon which a vibrant, healthy society depends. If anyone can believe anything for any reason, Clifford fears that ultimately the fabric of society will unravel because *a healthy society depends on people being able to trust others as a source of knowledge*. Credulous people cannot be trusted. Here's Clifford in his own words:

> Every time we let ourselves believe for unworthy reasons, we weaken our powers of self-control, of doubting, of judicially and fairly weighing evidence. We all suffer severely enough from the maintenance and support of false beliefs and the fatally wrong actions which they lead to, and the evil born when one such belief is entertained is great and wide. *But a greater and wider evil arises when the credulous character is maintained and supported*, when a habit of believing for unworthy reasons is fostered and made permanent. ...[I]f I let myself believe anything on insufficient evidence, *there may be no great harm done by the mere belief; it may be true after all, or I may never have occasion to exhibit it in outward acts. But I cannot help doing this great wrong towards Man, that I make myself credulous. The danger to society is not merely that it should believe wrong things, though that is great enough; but that it should become credulous, and lose the habit of testing things and inquiring into them; for then it must sink back into savagery*.

> The harm which is done by credulity in a man is not confined to the fostering of a credulous character in others, and consequent support of false beliefs. Habitual want of care about what I believe leads to habitual want of care in others about the truth of what is told to me. Men speak the truth to one another when each reveres the truth in his own mind and in the other's mind; but how shall my friend revere the truth in my mind when I myself am careless about it, when I believe things because I want to believe

them, and because they are comforting and pleasant? Will he not learn to cry, "Peace," to me, when there is no peace? *By such a course I shall surround myself with a thick atmosphere of falsehood and fraud, and in that I must live. It may matter little to me, in my cloud-castle of sweet illusions and darling lies; but it matters much to Man that I have made my neighbors ready to deceive.* The credulous man is father to the liar and the cheat; he lives in the bosom of this his family, and it is no marvel if he should become even as they are. So closely are our duties knit together, that whoso shall keep the whole law, and yet offend in one point, he is guilty of all.

Here is one way of reconstructing Clifford's argument:

1. If we believe things on insufficient evidence, we weaken our powers of being critical, doubting, and carefully weighing evidence.
2. Individuals' whose critical powers are weakened are more credulous.
3. When individuals are credulous they are easier to deceive and less trustworthy sources of information.
4. If individuals are easier to deceive, there will tend to be more deception.
5. If individuals are less trustworthy, then they will tend not to be trusted.
6. Therefore, a society in which individuals believe things on insufficient evidence would tend to be a society in which there is more deception and less trust. (from 1-5)
7. A society in which there is more deception and less trust is a worse society.
8. We ought not do things which will create a worse society.
9. Therefore, we ought not believe things on insufficient evidence. (from 6-8)

So, like the ship owner case, Clifford thinks that there is a general kind of harm (a "harm to Man") that is caused by believing things on insufficient evidence, *regardless of what the specific belief is*. The issue for Clifford is not the *truth* of the belief, but the *process* by which we come to believe it. He thinks that believing things too quickly and easily actively harms society. Clifford thinks we have an ethical duty to be critical of what we come to believe. In a famous passage Clifford notes,

> If a man, holding a belief which he was taught in childhood or persuaded of afterwards, keeps down and pushes away any doubts which arise about it in his mind, purposely avoids the reading of books and the company of men that call into question or discuss it, and regards as impious those questions which cannot easily be asked without disturbing it—the life of that man is one long sin against mankind.

How should we assess Clifford's argument? Is he right that there is something wrong with believing things we want to be true without looking carefully at the evidence? Are the results for society as dire as Clifford warns? One thing to be wary about in Clifford's argument is the possibility of a **slippery slope fallacy.** Note that premises 1-5 are all **conditional statements**—statements of the form "if A then B." The form of the first part of the argument goes like this:

> If A then B
> If B then C
> If C then D
> If D then E
> Therefore, if A then E

In logic, this is called a hypothetical syllogism; notice how each premise functions like a link in a chain where the conclusion links the first link (A) with the last link (E). However, notice the most plausible reading of those premises is not that every time x occurs, y occurs but that *most* of the times x occurs y occurs (this is clear from the language of "tends to" in some of the premises). In this case even if each of the five conditional statements were fairly probable (for example, B might follow from A 80% of the time), when we link them all together, the conclusion is no longer probable. This is because of how probability works: the probability of the conclusion decreases the more probabilistic conditional statements you add. Since the above reconstruction of Clifford's argument contains four probabilistic conditional statements, the likelihood of the conclusion is: .8 x .8 x .8 x .8 = .41. That means the conclusion is actually very *unlikely*! That is not a strong argument. This is one reason to doubt that Clifford's claims about the negative effects of wishful thinking on society are as bad as he worries they are.

One thing that bears on Clifford's argument and that was discovered long after Clifford was alive is what psychologists call "confirmation bias." **Confirmation bias** is a cluster of phenomena that all have in common the idea that we are more likely to believe something that we want/wish to be true.[4] What psychologists have shown over the last few decades is that individuals are less likely to pay attention to evidence that disconfirms something they believe than to evidence that confirms it. Moreover, individuals will tend to forget evidence that challenges an existing belief and more likely to remember evidence that supports it. This is exactly the kind of thing that Clifford was saying that we shouldn't do, but what psychologists have discovered is that we all in fact *do* do it. A liberal will exhibit confirmation bias in favor of her liberal beliefs; a conservative will exhibit confirmation bias in favor of her conservative beliefs.

---

4. For a scholarly review of the phenomenon of confirmation bias, see Raymond Nickerson, "Confirmation bias: A ubiquitous phenomenon in many guises" in *Review of General Psychology,* vol. 2, no. 2, pp. 175-220. The general idea of confirmation bias is not new, however. The philosopher Aurthur Shopenhauer anticipated it in 1819: "A hypothesis, conceived and formed, makes us lynx-eyed for everything that confirms it, and blind to everything that contradicts it. What is opposed to our party, our plan our wish, our hope often cannot possibly be grasped and comprehended by us, whereas it is clear to the eyes of everyone else" (*World as Will and Representation*, vol. 2).

And so on for every other social category we can imagine, including atheists! The facts that the trait that Clifford was warning against is so widespread might be taken to reinforce Clifford's argument. One way of objecting to Clifford's argument, for example, would be to say that he is just wrong about how credulous humans tend to be. But the evidence that psychologists have gathered over the decades on confirmation bias defuses this objection: we all have the tendency to believe what we want to believe. But, again, the question is whether this is such a bad thing?

Here is one reason for thinking that it *is* a bad thing and that we should, instead, foster exactly the kind of critical awareness that Clifford recommends. In an increasingly polarized era, where politics increasingly feels like a team sport and there is precious little quality conversation between groups that hold opposing views, one might worry about the health of the democracy. If we all surround ourselves with only those who tend to share our views, perhaps healthy political discourse will break down. Perhaps it already has. Political discourse seems not to be driven by evidence but, rather, by forgone conclusion that each side is to accept, regardless of evidence. If that is at all a somewhat accurate diagnosis of the current state of politics in the United States (and also elsewhere in the world), then perhaps fostering a more critical attitude towards one's own beliefs would be salutary. Maybe other domains of life, including religion, would also become less susceptible to confirmation bias and hence less polarized if we were to foster a more critical attitude towards our beliefs, tying them more strongly to evidence as Clifford suggests we ought to. The point of the foregoing is that given that we now know (scientifically) that human beings are naturally susceptible to the vices of credulity that Clifford warns against, perhaps this provides a further reason for thinking that Clifford was onto something, even if he sometimes comes off, in the words of William James, "with somewhat too much of robustious pathos in the voice."[5]

The will to believe

In his famous essay, "The Will to Believe," the American philosopher and psychologist, William James (brother of the famous novelist Henry James) argues against Clifford's "ethics of belief." James thinks that there are plenty of cases where it is not morally wrong to believe something without sufficient evidence. James takes himself to be arguing in support of Pascal's position that it is sometimes ok to believe certain things for pragmatic reasons. James thinks that pragmatic reasons are always context-specific—that is, they only work for beliefs that are already "live options" for us in our specific cultural contexts. But what determines which things are considered live options in a particular cultural context is almost never evidence, but other emotional and

---

5. This line is from William James's famous essay, "The Will to Believe" (section III) which will be discussed in the next section.

volitional (desire- based) factors. James makes this point with religious belief, specifically, which is the not-so-subtle subtext lurking behind the whole debate on the ethics of belief. Commenting on Pascal's wager, he says:

> You probably feel that when religious faith expresses itself thus, in the language of the gaming-table, it is put to its last trumps. Surely Pascal's own personal belief in masses and holy water had far other springs; and this celebrated page of his is but an argument for others, a last desperate snatch at a weapon against the hardness of the unbelieving heart. We feel that a faith in masses and holy water adopted willfully after such a mechanical calculation would lack the inner soul of faith's reality; and if we were ourselves in the place of the Deity, we should probably take particular pleasure in cutting off believers of this pattern from their infinite reward. It is evident that unless there be some pre-existing tendency to believe in masses and holy water, the option offered to the will by Pascal is not a living option. Certainly no Turk ever took to masses and holy water on its account; and even to us Protestants these means of salvation seem such foregone impossibilities that Pascal's logic, invoked for them specifically, leaves us unmoved. As well might the Mahdi write to us, saying, "I am the Expected One whom God has created in his effulgence. You shall be infinitely happy if you confess me; otherwise you shall be cut off from the light of the sun. Weigh, then, your infinite gain if I am genuine against your finite sacrifice if I am not!" His logic would be that of Pascal; but he would vainly use it on us, for the hypothesis he offers us is dead. No tendency to act on it exists in us to any degree.

James's point is that pragmatic arguments only work for beliefs that are already live options for us. But how those beliefs come to be live options in the first place, is rarely a matter of evidence:

> It is only our already dead hypotheses that our willing nature is unable to bring to life again. But what has made them dead for us is for the most part a previous action of our willing nature of an antagonistic kind. When I say 'willing nature,' I do not mean only such deliberate volitions as may have set up habits of belief that we cannot now escape from,—I mean all such factors of belief as fear and hope, prejudice and passion, imitation and partisanship, the circumpressure of our caste and set. As a matter of fact we find ourselves believing, we hardly know how or why. Mr. Balfour gives the name of 'authority' to all those influences, born of the intellectual climate, that make hypotheses possible or impossible for us, alive or dead. Here in this room, we all of us believe in molecules and the conservation of energy, in democracy and necessary progress, in Protestant Christianity and the duty of fighting for 'the doctrine of the immortal Monroe,' all for no reasons worthy of the name. We see into these matters with no more inner clearness, and probably with much less, than any disbeliever in them might possess. His unconventionality would probably have some grounds to show for its conclusions; but for us, not insight, but the prestige of the opinions, is what makes the spark shoot from them and light up our sleeping magazines of faith. Our reason is quite satisfied, in nine hundred and ninety-nine cases out of every thousand of us, if it can find a few arguments that will do to recite in case our credulity is criticized by some one else. Our faith is faith in some one else's faith, and in the greatest matters this is most the

case. Our belief in truth itself, for instance, that there is a truth, and that our minds and it are made for each other,—what is it but a passionate affirmation of desire, in which our social system backs us up? We want to have a truth; we want to believe that our experiments and studies and discussions must put us in a continually better and better position towards it; and on this line we agree to fight out our thinking lives. But if a pyrrhonistic skeptic asks us how we know all this, can our logic find a reply? No! certainly it cannot. It is just one volition against another,—we willing to go in for life upon a trust or assumption which he, for his part, does not care to make.

James is treading deep philosophical waters here at the end of this passage. He is noting that many of the things we believe, including things like that there is an external world at all, are not underwritten by evidence-based reasons. How can you defeat the skeptic with evidence, after all?[6] Or consider the idea that a good society is one that protects those who are vulnerable and not capable of

protecting themselves. Can you prove that with reasons or evidence? It seems not since it is such a basic and obvious point to us—not in need of further justification. James's point is that there are many types of beliefs that we do not have sufficient evidence for and religious belief is one of them. In cases like these, we do nothing morally wrong in holding onto the belief, contrary to what Clifford claims.

As James sees it, Clifford is mistaken in thinking that there is something morally wrong with believing things without sufficient evidence. The specific beliefs that James focuses on are moral beliefs, interpersonal relationship beliefs, and religious beliefs. James has us consider two belief-forming policies:

- <u>Clifford's policy</u>: believe only that for which you have sufficient evidence
- <u>James's policy:</u> believe live options when they are important for your life, even if you don't have sufficient evidence for them

James claims that although following Clifford's policy would tend to result in having true beliefs (because it's a stricter policy), it would also potentially cause you to miss out on some truths that didn't reveal themselves so easily. Suppose the universe were such that there was a god, but that god didn't reveal themself unless people were willing to take a leap of faith in the first place. If that were so, then someone who followed Clifford's policy would never be able to acquire the truth about god's existence. Here is James:

We feel, too, as if the appeal of religion to us were made to our own active good-will, as if evidence might be forever withheld from us unless we met the hypothesis half-way. To take a trivial illustration: just as a man who in a company of gentlemen made no advances, asked a warrant for every concession, and believed no one's word without proof, would cut himself off by such churlishness from all the social

---

6. For more on external world skepticism, see that chapter in this textbook.

rewards that a more trusting spirit would earn,—so here, one who should shut himself up in snarling logicality and try to make the gods extort his recognition willy-nilly, or not get it at all, might cut himself off forever from his only opportunity of making the gods' acquaintance. ... I, therefore, for one cannot see my way to accepting the agnostic rules for truth-seeking, or willfully agree to keep my willing nature out of the game. I cannot do so for this plain reason, that *a rule of thinking which would absolutely prevent me from acknowledging certain kinds of truth if those kinds of truth were really there, would be an irrational rule*. That for me is the long and short of the formal logic of the situation, no matter what the kinds of truth might materially be.

James's argument in "The Will to Believe" thus consists of two separate lines of reasoning, both of which we have already discussed. First, since many of our beliefs are such that they cannot be based *wholly* on sufficient evidence (since the winnowing down of beliefs to those we see as live options is are driven by cultural forces over which individuals have little control), Clifford's policy would be unduly restrictive and unfair (and we ought not endorse policies that are unfair). Second, Clifford's policy would turn us all into skeptics about the important matters of life and we are not morally obligated to be skeptics (being a non-skeptic is not more immoral than being a skeptic). As James notes at the end of "The Will to Believe,"

> In all important transactions of life we have to take a leap in the dark... If we decide to leave the riddles unanswered, that is a choice; if we waver in our answer, that, too, is a choice: but whatever choice we make, we make it at our peril. If a man chooses to turn his back altogether on God and the future, no one can prevent him; no one can show beyond reasonable doubt that he is mistaken. If a man thinks otherwise and acts as he thinks, I do not see that any one can prove that *he* is mistaken. Each must act as he thinks best; and if he is wrong, so much the worse for him.

It is important to note that James has not so much rebutted Clifford's moral argument as he has given a different argument with a different conclusion. That is, James nowhere discusses the logic of Clifford's moral argument that was laid out above. Rather, he provides a different argument against Clifford's conclusion. There's nothing wrong with that tactic, but it does leave one wondering how Clifford's argument can be criticized. As often happens in philosophy, we are left with the raw materials to construct our own view on the matter. Both Clifford and James have given interesting arguments for differing conclusions; our task is to knit those somewhat unconnected pieces into a satisfying view.

What might Clifford say in response to James's argument? One thing I think he might say is that even if it is true that no one's beliefs are ever totally evidence-based and bias free, it doesn't follow that we can't seek to do better. Clifford is clearly not as worried about having false beliefs as he is with a moral duty to consider evidence over wishful thinking. Even if it is true that it is very difficult to eliminate all wishful thinking, as the phenomenon of confirmation bias suggests, that doesn't mean that we can't at least try to do better. On this point, perhaps James and Clifford are in agreement. James is not arguing that anyone can believe anything they wish, but that with certain beliefs we cannot base them wholly on evidence. And it seems that Clifford would

agree here: even if I cannot refute the moral skeptic, it doesn't follow that the moral beliefs I have are purely wishful thinking. What about religious belief—what would Clifford say about that? Suppose it turned out (as in James's scenario above) that there were a god but that god was elusive and didn't reveal themself to those who didn't take some steps of blind faith. Presumably Clifford's tendency would be not to believe in god based on the lack of evidence. If Clifford were to find himself before this god at some point, perhaps he would reply, "Sir why did you not give me better evidence?"[7]

Study questions

1. True or false: Pascal thinks that it is very likely that god exists.
2. True or false: Pascal's wager is a pragmatic argument
3. True or false: Pragmatic arguments are attempts to get something to believe x, regardless of whether x is true or if there's good evidence to believe x.
4. True or false: Pascal's wager is one of the earliest examples of decision theory.
5. True or false: Clifford thinks that false beliefs are morally wrong.
6. True or false: Clifford thinks that what makes the ship owner's action wrong is that his false belief caused others harm.
7. True or false: Clifford thinks that as long as what I believe doesn't directly harm anyone, then I should be allowed to believe it—*even if* it is not supported by sufficient evidence.
8. True or false: What psychologists call "confirmation bias" is exactly the kind of thing that Clifford railed against—a kind of wishful thinking (or believing).
9. True or false: James thinks that most of our beliefs are driven *only* by evidence.
10. True or false: James thinks that if we were to follow Clifford's policy then it would turn us into skeptics—specifically skeptics about moral and religious truths.
11. True or false: Clifford thinks that it is permissible to believe some things that aren't based on sufficient evidence.
12. True or false: James thinks that it is permissible to believe some things that aren't based on sufficient evidence.

---

7. This is actually a famous response from the philosopher Bertrand Russell who was once asked in an interview in the *Saturday Review* (1974): "Let us suppose, sir, that after you have left this sorry vale, you actually found yourself in heaven, standing before the Throne. There, in all his glory, sat the Lord—not Lord Russell, sir: God." Russell winced. "What would you think?" "I would think I was dreaming." "But suppose you realized you were not? Suppose that there, before your very eyes, beyond a shadow of a doubt, was God. What would you say?" The pixie wrinkled his nose. "I probably would ask, 'Sir, why did you not give me better evidence?' "

13. True or false: Pascal thinks that it is permissible to believe some things that aren't based on sufficient evidence.

For deeper thought

1. In William James's "Will to Believe," James construes Clifford's position as *being driven by the fear of being wrong* (that is, by the fear of having false beliefs). James points out that the fear of being wrong is itself something "passional" (meaning "emotional") and thus not purely rational or evidential. Thus, he thinks Clifford is being hypocritical in a certain sort of way. Do you think that James has accurately captured Clifford's view, based on what was presented above, or has James committed a straw man fallacy? Why or why not?

2. What do you think about James's idea that there are certain things in life that require a leap of faith before you can see them to be true? Are there things in life like this? If so, how do you think Clifford would respond? Would he think that it is wrong to pursue something that we think might be true in order to see if it is true?

3. Consider the examples that William James puts forward in section IX of "The Will to Believe." Do you think that all of these examples are cases in which the individuals need to *believe* x is true or do they only need to *hope* x is true?

4. Why does Clifford think that it is immoral to believe on insufficient evidence even when no one will be directly harmed by those beliefs?

CHAPTER 8: HUMAN WELL-BEING

HUMAN WELL-BEING

Matthew Van Cleave

How do I live well? What is needed in order for my life to go well for me? In answering this question, many people would start listing different kinds of things that they desire and whose possession they believe would make their lives better. A common response for many people living in the United States today would probably be "having more money." There's something right about this response since it is difficult for us to imagine someone who is constantly having to worry about making ends meet as living well. But suppose that one lived in a utopian communist society where there wasn't money and all one's needs were met without the use of money—in such a society, would having money make one's life better? Clearly not. So perhaps our original question can only be answered relative to a particular cultural context. This much seems to be obviously true. However, is there a way of answering our original question that is more universal? Is there any way to say *in general* what humans need in order to live well? If someone's life is going well then what is it that that person *has* and that other people (whose lives are not going well for them) *lack*? These questions are addressing what philosophers call "**well-being**" and I will refer to this as the **question of well-being.** The question of well-being is a very old one that can be traced all the way back to Socrates in ancient Athens (469-399 BCE). In this chapter we will consider two competing ways of answering the question of well-being. **Subjectivist theories of well-being** claim that well-being is nothing other than the *satisfaction of one's desires* (whatever they are). In contrast, **objectivist theories of well-being** claim that there are certain *specific things* that are needed in order to attain well-being. There could be many different versions of the objectivist theory. For example, one might claim that human well-being requires good health, good morals and happiness. To lack any one of these things, would be to lack the full measure of well-being that is possible for human beings. A different objectivist theory might put different things in the list, perhaps leaving some of those things out but adding others: autonomy, happiness, and friendship, for example. In the following we will explore the pros and cons of these different theories—the arguments for and against them—and see what kinds of things are at stake in this age-old debate.

There are some things that we value because they enable us to get something *else* and other things that we value for their own sake. A good example of the former would be money: the reason we value money is because of the other things it can get us (a house, a car, healthcare, security, and so on). Another good example of something we value because it enables us to get something else is gasoline (if we have a car, lawnmower, or something that uses gasoline). If one didn't have a car or lawnmower then gasoline would no longer have any value. Philosophers refer to things that have value for us only because we value something else **"instrumentally valuable."** Gasoline and money are paradigm cases of things that have instrumental value because they do not directly improve our well-being but do so only because they enable us to get *some else*. In contrast, things that

we value directly, for their own sake and not for any other reason, are said to be "**intrinsically valuable.**" Some examples of things that have intrinsic value are friends, family members, a child, or your beloved pet.

We value these things in and of themselves and not simply because of what they bring us. Even if your child is making you very angry, you still love and value them. Another way of reformulating the question of well-being is as a question about intrinsic value: what things in the world have intrinsically value? The subjectivist's answer to this question is: only the things we desire. The objectivist's answer will differ depending on what version we are considering, but fairly common to any version of the objective theory will be things like *happiness* and *autonomy*. It is characteristic of the objective theory to claim that if something is truly intrinsically valuable then it remains so even if a person does not desire that thing. Thus, for example, the objectivist would say that happiness makes a person's life better (increases their wellbeing) even if that person doesn't desire to be happy.

Subjectivist theories of well-being

Subjectivist theories of well-being have a lot going for them. To see why, consider the following question: who is the person who is best placed to be able to determine whether my life is going well or not? The subjectivist's answer to this is that *I* am the person who can best determine whether my life is going well—not my parents, my pastor, my teachers, my friends, or anyone else.

Perhaps my parents think it would be best for me to become a banker or my pastor thinks it would be best for me to go to church. The subjectivist denies that these really are good for me if I do not desire them. No one else but me can determine what is best for me—that is the subjectivist's view on the matter. And that seems to accord with common sense since we believe ourselves to be the authorities on *whether* our lives are going well for us. How could anyone other than ourselves possibly determine that? We can thus say that subjectivist theories accord individuals a high level of **personal authority**. It seems to follow from this that people cannot really be mistaken about whether their lives are going well for them. That is, according to a subjectivist, if one thinks that one's life is going well then it is going well and if one thinks that it is going poorly then it is going poorly. Perhaps sometimes people don't know why their lives are going poorly and thus they seek out the help of a psychologist or other professional. But seeking help in this way is consistent with what the subjectivist is claiming since the person seeking the help knows *that* their life isn't going well (they just need help in figuring out *why*). It's not as if one would ever go to see a psychologist or other professional only to be told that actually their life was perfectly fine and then accept that they must have been mistaken. For the subjectivist, well-being is like physical pain. You can't be wrong about whether you are in pain: if you think you're in pain, then you are in pain. Likewise, according to the subjectivist, if you think your life is going well, then it is (and vice versa). Thus we can say that according to the subjectivist, our own assessment of our well-being is **infallible.**

What most sets subjectivist theories apart from objectivist theories is this issue of fallibility. For the subjectivist,

our well-being depends on whether or not our desires are satisfied—whether or not we get what we want. But those desires themselves cannot be criticized. If what I most want is to be a hermit and live in the mountains herding sheep, then achieving this is what would make my well- being high. If what I want is to play video games all day while eating junk food, then this is what would make my well-being high. I might not always know what it is that I most deeply desire—perhaps I need help figuring out what my deepest desires are—but whatever they are, satisfying them is what would be most conducive to my well-being. For the subjectivist, there's no sense in which my desires themselves can be flawed or mistaken or not good for us to have.

Rather, my desires *define* what is good for me. An apparent exception to this would be cases in which some of my desire conflict with other desires. For example, suppose that I really want to each chocolate cake but I am allergic to chocolate. In that case, perhaps it wouldn't be good for me to eat the chocolate cake even if I want it. However, this is only because there is some *other* desire that is stronger—in this case, the desire not to have an allergic reaction to the chocolate cake. But there's no reason that the subjectivist can't account for what's going on here. It is simply a case in which one desire (the desire not to have an allergic reaction) is stronger than the other (the desire to eat the cake) and in such cases the subjectivist claims that what is in our best interest is to act on the strongest desire. So the only reason for not acting on a desire is because doing so would undermine an even stronger desire. This would be the subjectivist's way of explaining why one shouldn't act on certain strong desires if those desires are illegal and one had a high chance of getting caught. For example, a serial killer who strongly desired to kill people (think of someone like Hannibal Lecter) would have their well-being increased by killing people. The only reason that acting on these desires would be bad for Hannibal is because of the possibility that he would get caught and go to prison (and thus no longer be able to live his serial killer life anymore). If Hannibal also strongly desired to not get caught and go to prison, then it is that stronger desire that would act as a constraint on his desire to kill people in creative ways. But the desires in and of themselves are beyond criticism, according to the subjectivist. If we imagine a world in which Hannibal knew for sure he would never get caught, then the subjectivist would admit that killing people unequivocally increases Hannibal's well-being. The same point applies to less macabre cases, such as the drug addict who wants nothing other than to be high. If that is what they most desire, then the satisfaction of that desire unequivocally increases that person's well-being. Imagine the son of a wealthy but negligent parent who funds their son's drug habit, paying for all of their living expenses and drugs. If all the son wanted to do in life was drugs (and nothing else), then the satisfaction of this desire is what would most increase his well-being. This follows directly from subjectivism's features of personal authority and infallibility, outlined above.

Since everyone's desires are different and since subjectivism defines an individual's well-being in terms of the individual's desires, it follows that well- being is pluralistic for the subjectivist. **Pluralism** here simply means that there are many different kinds of lives that would count as equally well-lived. Since the only thing that has intrinsic value for the subjectivist is one's deepest desires, and since individuals' desires can differ radically, it follows that what it means for individuals to have a high level of well-being can also differ radically. The drug addict son of the wealthy negligent parent would have a high level of well-being and so would the self-taught

musician who works hard throughout their life to finally achieve widespread acclaim and a Grammy award and so would innumerable other kinds of lives: the fireman who wanted nothing more than to be a fireman his whole life, the physicist who wanted nothing other than to be a physicist her whole life, the lawyer, the philosopher, the triage nurse working in a war zone, the stay-at-home father, the photojournalist documenting the atrocities of war. All of these very different kinds of lives would count as equally well-lived *as long as they had achieved what they most deeply desired*. If we think that there are indeed many, many different ways of living equally well-lived lives, then subjectivism explains why this is so.

Objectivist theories of well-being

As noted above, what most distinguishes objectivist from subjectivist theories is their stance on the relationship between desires and well-being: subjectivists think that our desires define well-being and thus do not think our desires can be flawed or mistaken whereas objectivists think that our desires *can* be flawed or mistaken. The most minimalist type of objectivism is a view called **hedonism**.

Hedonists say that there's only one thing needed for well-being and that is happiness. It doesn't matter what makes you happy; all that matters is that you are happy. As long as you are happy, the hedonist says, your life is going well for you. On the other hand, if you are not happy at all—in the extreme, if you have never been happy—then your life is going maximally poorly for you.

Hedonists also conceive of happiness not simply as **sensate pleasure** (think of things like the taste of really good food, the feeling of sex, or a really nice back massage), but as **enjoyment.** Enjoyment is different than sensate pleasure because it is a state of mind or attitude that you take towards some experience or state of affairs. For example, you might enjoy the feeling in your stomach when you go down the big hill in a roller coaster or you might enjoy certain kinds of painful sensations, such as having a really tight muscle massaged. This is why a masochist (a person who enjoys inflicting pain on themselves) is also being a good hedonist: they enjoy the pain. For the hedonist, happiness is about enjoyment and different people can enjoy *very* different kinds of things.

What makes hedonism an objectivist rather than subjectivist account of well- being is that people might not want to be happy. Suppose someone sincerely desires to be unhappy—and not just for a while, but all the time. Suppose they desire this because they think that it is proper or appropriate for them to be unhappy. Perhaps they even think that their well-being consists in their being unhappy. Can you think of such a case? (Stop and actually try to come up with a case before reading on.[1]) If there are such cases, then the difference between

---

1. If you've thought about it and can't come up with a case, here are two cases to consider. Case 1: the religious ascetic. Imagine a person who believes that their earthly existence should be characterized by pain and suffering and that only by

hedonism and subjectivism becomes clear. Hedonism would say that the lives of such individuals are not going well, since they are not happy. Subjectivism, on the other hand, would say that insofar as unhappiness is what such individuals most deeply desired, it follows that their well-being was high. This might sound paradoxical, but it follows from defining well-being in terms of what an individual most deeply desires.

Nevertheless, hedonism is a very subjective kind of objectivism. What I mean is that it allows for a very wide range of different kinds of lives that count as equally well-lived—what we have called pluralism. It is not completely pluralistic, since it would say that someone who desired to be unhappy (and achieved that desire) as having low well-being. Nevertheless, it would still say that the serial killer and rich, lazy drug addict had high well-being. However, these kinds of cases might seem to pose an objection to hedonism: if hedonism says that these individuals lives are going really well for them and if we think that these aren't really good lives that are being lived, then perhaps that means there is something wrong or limited about hedonism as a theory of well-being. It is for this reason that objectivists typically move beyond hedonism, adding other things besides happiness, that are required for a life well-lived. What other things might be required for well-being besides happiness? Some common things that philosophers suggest are: **autonomy, personal growth, friendship, morality,** and **knowledge**. Consider what it would mean to say that personal growth is necessary for one to achieve a well-lived life. If personal growth is necessary, then the life of the rich, lazy drug addict who did nothing other than sit around all day getting high would not count as a well-lived life, *even if we grant that that person enjoys this life*.

You might think that some of the things listed correlate well with happiness. For example, if people have more autonomy in their lives, they tend to be happier. It is an interesting empirical question whether this is so, but even if it is, we can nevertheless envision lives that have the highest levels of happiness but that lack autonomy. For example, Aldous Huxley's *Brave New World* presents a society in which everyone is happy, but they have been (in essence) conditioned to be happy. For example, citizens are conditioned to take a special drug called "soma" any time they begin to feel unhappy. The drug makes them feel happy again and they don't even really understand the ways they've been conditioned by the controllers of the society. Moreover, they've been conditioned from a young age to only want the things that the controllers of the society deems it legitimate for

---

living in such a way will they achieve redemption from their sins. There were a number of such individuals within the Eastern Orthodox Christian tradition. It seems that such individuals believed that their earthly well-being consisted in their suffering, not in their happiness. Case 2: a person with guilty conscience. Consider a person who had done something really horrible, for example, they had accidentally run over their 3 year old daughter while backing out of the driveway while intoxicated. Or a soldier who had had to leave behind a wounded friend on the battlefield to die. Such individuals might feel that they do not deserve to be happy. They might sincerely believe that the way they ought to live is in a state of perpetual unhappiness. To the extent that such an individual began to experience any level of happiness, they might experience deep guilt and thus try to make themselves unhappy again.

them to want. In this way, the society has been set up so that everyone always gets what they want (and so are almost always happy) and in the few cases in which they aren't happy, they take a magical wonder drug that makes them happy again. The lives of such individuals certainly contain happiness, but they do not contain autonomy since they are being controlled by external factors that they do not understand. For this reason, objectivists who think that autonomy is necessary for well-being would deny that the lives of the citizens of *Brave New World* have the highest level of well-being, even if we grant that they are as happy as possible. Since the people in Brave New World are not in control of their lives, they lack autonomy and thus are not living as well as they could be.

One of the points in favor of the objectivist account of well-being is that it allows us to say what is deficient about certain kinds of lives that we do not consider to be well-lived, even if the person living them thinks they are. The objectivist for whom autonomy is necessary for living well can explain what is lacking in the lives of the citizens of Brave New World. The objectivist for whom person growth is necessary for living well can explain what is lacking in the life of the rich, lazy drug addict. The objectivist for whom morality is necessary can explain what is deficient in the life of the serial killer. And so on.

Is morality necessary for living well?

Jake is a morally despicable person. He lives selfishly without regard for others and regularly lies, cheats and steals. In general his is willing to harm other people in order to get what he wants. After enough time, Jake eventually has lots of money and cushy life but he has no friends. Or at least no true friends. The "friends" he has are people whom he pays, in some form or other, to be around. Jake is truly alone. But he doesn't care; he likes it that way. Is Jake living well? He is according to a subjectivist account of well-being, but he wouldn't be according to an objectivist account that requires any of the items listed above. In particular, Jake's life lacks moral goodness and so any account of well-being that required that would count Jake's life as not as well-lived as it could have been had he been a moral person. Of course, Jake's life also lacks friendship and arguable he lacks friendship *because* he lacks morality. If so, it appears that morality is functioning as something that has instrumental value: morality is important because without it one cannot have significant friendships. Friendship, in this case, is the thing that has intrinsic value and morality is important only because it helps us to achieve it. If this is so, the objectivist would not say that morality itself is necessary for a good life, since its value is only instrumental, not intrinsic.

So is morality *itself* necessary for living well, apart from any other things of value which it might bring about? Plato famously thought that morality (justice) was necessary for living well and he defends a view like this in his most famous work, *The Republic*. However, contrary to what Plato thought, there are good reasons for thinking that morality is not necessary for human well-being. Imagine a character who cheats his way to the top, trampling over others to get there.

This could be an investment banker, real estate tycoon, or even the leader of nation. By all appearances from the outside, this person is happy, has significant relationships, possesses autonomy and personal growth. It really seems that this person has it all, it's just that they are a total asshole. Why do we resent such a person? Arguable we resent them because their life is going really well despite the fact that they are not a morally good person. If it were a really bad person whose life was also not going well for them then we might not resent them so much. It is the fact that this person's life is going well that makes us resent them. It seems unfair and unjust that they should be living well while being so despicable. But notice that in order to make sense of our resentment in this case, we are granting that this person's well-being is very high, which seems to entail that morality is *not* necessary for a high level of well-being. Plato thought that if an individual was immoral this was evidence of a soul in disarray—of someone who lacked "psychic harmony." Such a person, Plato thought, was really unhappy on the inside, contrary to external appearances. Thus Plato thought that we shouldn't really envy or resent such a person since, contrary to appearances, their well-being wasn't really high at all. This is his answer to the question of why we ought to be moral: because to be moral is to be in a state of psychic harmony and the state of psychic harmony *defines* well-being. But how could Plato *know* that psychic harmony defines well-being? In the case of an individual who outwardly appeared to be living as well as possible, why think that they weren't? Plato's theory seems to run counter to the empirical evidence, which is that there are cases of people who live highly immoral lives and yet have the highest levels of well-being by any way we could measure.

Return to the case of the serial killer who is living well according to any of the measures we have considered: he is happy, autonomous, has achieved personal growth, is knowledgeable, and so on. If you doubt this could be the case then just consider an individual like Hannibal Lecter. Hannibal is certainly not living a moral life, but his life certainly seems to be going well for him—that is, he has a high level of well-being. The question of well-being is not a question about morality. Highly immoral people are able to live lives that go well for them by any of the measures we have considered. To claim that morality is non- instrumentally necessary for well-being seems to go against the evidence. One can live well without living morally. That is why we hold such resentment towards such people.

Aristotelian accounts of human well-being

One way of arguing for objectivist accounts of well-being is to determine what kinds of things human beings universally, cross-culturally value. If a kind of life that lacks autonomy is universally judged to be not as good as one that has autonomy then *ceteris paribus* (Latin for "all other things being equal") a life that has autonomy is better than a life that lacks it. The same goes for other things that objectivists think directly (non-instrumentally) improve one's well-being— happiness, growth, knowledge, and so on. In short, we know that these things are intrinsically valuable because people have always (cross-culturally) valued them. This raises the question of *why* it is that people value these things but it also raises the question of whether people are *right* to. After all, for most of human history human societies have practiced various forms of domination and

subjugation of the weak and disenfranchised but this does not mean that they were *right* to do so. One way of trying to answer these questions is to advert to claims about human nature. What I am calling an **Aristotelian account of well-being** is an objectivist account of well-being that grounds claims about well-being in claims about human nature. The key idea coming from Aristotle is the idea that what it means to say that something is a good x derives from facts about x's purpose. Aristotle's argument appears in a passage from a text called the *Nicomachean Ethics*. There Aristotle argues that the function or purpose of human beings is what only human beings can do. So, for example, the purpose of human life cannot simply be living, since that is something that humans share with plants. Nor can the purpose of human life be the mere use of our sensation (such as pain or pleasure), sense perception and locomotion, since other animals share those capacities with humans. So which of our capacities are characteristically human—ones that only human have? For Aristotle, it is our reason that is distinctively human. Since reason is our characteristically human capacity, it follows that human well-being requires developing excellence in those activities that utilize our reason. Without the development of our characteristically human capacities, humans could not achieve **eudaimonia**—the Greek word Aristotle used to capture the notion of human flourishing.

We needn't have a narrow view of what Aristotle means by "reason" and many philosophers who followed in Aristotle's footsteps didn't. Karl Marx, for example, uses an Aristotelian account of human well-being to ground his criticism of late 19th century industrial capitalism. By that time, factories had come to replace medieval guilds—groups of craftsmen—who would build, say, a chair or table from start to finish. Whereas craftsmanship took fine motor skills, thought, knowledge, and planning—all the kinds of things that fall under this broader notion of "reason," factory work was mind-dumbingly dull. A life that consisted of the same simple and thoughtless tasks performed over and over again did not enable individuals to develop their characteristically human capacities. Sure, they enabled wealthy business owners to increase their profits, but Marx saw that this was at the cost of the quality of the individual workers' lives.

One of the things that characterizes human beings is our use of culture to radically transform our basic biological activities. For example, both humans and non-human animals engage in activities like eating, mating, and fighting— these are essential biological activities for animals. But only humans transform these activities into something artistic—where the activities and associated objects themselves become ends in themselves rather than as a means to an end. From the biological perspective, the point of eating is to maintain the metabolism needed to live, the point of sex is to pass on one's genes, and the point of fighting is to establish dominance or survive violence directed at you. Only humans radically transform these activities into things where the purpose of the activity does not have to do with survival but rather with pleasure. We develop gourmet cooking where the main goal is not sustenance but development of our palates; we develop erotic literature, new sexual positions, and birth control, whose main goals are not reproduction but pleasure; and we develop things like mixed martial arts fighting whose main goal is not dominance or survival but development and demonstration of the art and skill itself. The main point of these activities is not their original biological function.

If you need to convince yourself of this just consider activities like wine tasting where you don't actually drink the wine, or a cooking show where the judges eat only a sample of the whole dish, or a boxing match between two friends or brothers. In short, humans create new domains with radically different purposes—purposes that transcend our narrow biological imperatives and thereby create new domains of value. This transcending of the merely biological can take place only insofar as we our using our characteristically human capacities.

To see how the Aristotelian account of human well-being compares to the subjectivist account of well-being, consider our earlier example of the wealthy, lazy drug addict. Assuming this person wants nothing other than to be high on drugs all day, the subjectivist would say that their well-being was as high as it could possibly be. But the Aristotelian would deny that this person's well-being was high at all because they are in no way using or developing their characteristically human capacities. On the other hand, in being a successful serial killer, the serial killer may be fully utilizing and developing their characteristically human capacities. If you doubt this then consider a character like Hannibal Lecter, a highly intelligent doctor and aesthete who seems to have developed his characteristically human capacities about as highly as any individual could. So in this case the Aristotelian and subjectivist accounts of well-being would actually agree.

The main problem with Aristotelian accounts of human well-being is that they attempt to derive an "ought" from an "is." In philosophy, the **is/ought gap** is the idea that one cannot logically derive a claim about what *ought to be* the case from a claim about what *is* the case. Claims about how humans ought to live cannot be simply derived from claims about how humans do live. The simplest way for me to explain the is/ought gap has to do with human reproduction.

1. Human beings have always reproduced
2. Therefore, human beings ought to reproduce

Does 2 follow from 1 alone? That is, if 1 is true, does 2 have to be true? To answer affirmatively is to attempt to bridge the is/ought gap. And that doesn't work. Just because human beings have always reproduced, it doesn't follow that we should continue to. That would only follow if we assumed that human ought to continue to exist and this is a separate "ought" claim that has been challenged by some![2] In general, just because x is (and has always been) the case, it doesn't follow from that alone that x ought to be the case. The "ought" claim is a new type of claim that can only be justified by some further "ought" claim.

The problem with Aristotelian accounts of human nature is that they make this same mistake in trying to derive a claim about what constitutes human well-being (how humans *ought* to live) from claims about human nature (how human beings *are*). Even we set aside questions about whether there really is a human nature,

---

2. A good example of this is the Voluntary Human Extinction Movement.

which some philosophical traditions (such as **existentialism**) have denied, there remain deep questions about how those claims are supposed to inform our account of human well-being. The subjectivist's retort to the Aristotelian is that if a person in no way desires x, then x will not improve that person's life at all. Consider a person who has tried developing their characteristically human capacities and doesn't like it at all. They don't care for gourmet food or pioneering new sexual positions or learning new things about the world. All they want to do is eat junk food (which has been designed specifically to target our more "animal" nature—fatty, salty, sugary foods) and watch reruns of Happy Days while sitting in a massage chair and having mood enhancing drugs pumped intravenously into their bodies. This is what makes this person happy and this is what they most want—nothing else is as pleasurable and everything else they find boring. The objection to the Aristotelian account of well-being is that it claims that this person's well-being would be increased precisely by not doing the things that they most want to do. And that seems not only heavy- handed but also false to subjectivists.

What's the meaning of life?

One question that might be thought to bear on the question of well-being is the question: what is the meaning of life? If there was a meaning of life, then it would seem that this should inform how we ought to live our lives. When philosophers ask if life has a meaning, by "meaning" they mean "purpose." Consider an individual's life: they live, performing many of the same activities day after day, and eventually die. One can pan out and consider not only individual human lives but humanity as such. At some point in the evolutionary past of the earth, early hominids evolved into *homo sapiens*—human beings.

Eventually humans developed the ability to ask themselves philosophical questions about the meaning of life. Many of our earliest myths can be seen as attempts to address questions about our purpose: where we came from, who we are, and what our destiny ultimately is. Are human beings simply a cosmic accident or is there a reason we are here—something we are supposed to accomplish as individuals, as a species? If there is a reason that human beings are here then knowing that reason would seem to be relevant to human well-being.

There are broadly two different views of the meaning of human existence. On one view, the only meaning that exists is the meaning that individuals can create for themselves in this lifetime. Call this **subjective meaning.** Subjective meaning could be things like having children or travelling or finding a career that one really loves. It is characteristic of subjective meanings that then end when that individual's life ends. Thus, subjective meaning is not permanent. Whatever meaning an individual was able to achieve in their life will pass away. In all likelihood, at some point in time any memories or knowledge of you will fade from this earth. We can put this picturesquely as saying that there is some point in time at which your name will be uttered for the last time. In contrast, **transcendent meaning** is a meaning that extends beyond your life here on earth—and even beyond the existence of the earth itself—and is permanent in the sense that it does not cease to exist. Allow

me to indulge for a minute and explain a specific view according to which life has transcendent meaning. I grew up in an Evangelical Christian family. It was a wonderful childhood. I had loving parents, wonderful siblings, was never in want, and had an excellent education with lots of interesting experiences. My family and I believed that this life and this earth would eventually come to an end, but that for those who believed the right things and lived the right way, they would continue to live a life of bliss in heaven. For my family growing up, life had a transcendent meaning: our purpose in life was to come to know god and to lead others to know god as well. The purpose of life on earth was to fulfill the "great commission" which basically meant convert the world's population to Christianity. The things that we fundamentally valued the most would continue to exist in another realm— heaven—for all eternity. According to this view, what it meant to live well was directly related to transcendent meaning. Such a view does not deny that subjective meanings exist, it just claims that they are not as important as transcendent meaning for the purposes of how we ought to live.

The most common way to account for the existence of transcendent meaning is through certain religious views of the world (such as the one explained above). That is because god/heaven is transcendent—that is, god/ heaven exists beyond this universe—and thus can ground transcendent meaning. In contrast to such views, naturalism denies that there exists anything beyond the universe that could ground transcendent meaning. According to naturalism[3], which draws on a scientific understanding of the universe (including of human beings), there is only subjective meaning. All of human existence is just a temporary blip in the history of the universe, for the naturalist. In the end, humans will no longer exist, nor will the earth and nothing that humans have done will have ever had any effect on the cold, impersonal universe. Human beings are here for a short time in the Earth's history and the Earth itself is just a minor speck in a vast universe. In the end, the universe itself will probably cease to exist, which is what many cosmologists predict based on the fact that the universe is expanding. And what will it all have been for? The naturalist's answer is: nothing. Human beings and the universe itself are just a cosmic accident; there is no purpose of it all.

According to the philosopher Richard Taylor, a long, drawn-out process which achieves nothing lasting in the end is the essence of meaninglessness. He gives a picturesque example:

> [T]here are caves in New Zealand, deep and dark, whose floors are quiet pools and whose walls and ceilings are covered with soft light. As you gaze in wonder in the stillness of these caves it seems that the Creator has reproduced there in microcosm the heavens themselves, until you scarcely remember the enclosing presence of the walls. As you look more closely, however, the scene is explained. Each dot of light identifies an ugly worm, whose luminous tail is meant to attract insects from the surrounding

---

3. "Naturalism" is a term that has many, many meanings in philosophy. As I am using the term here, it simply means the denial of supernaturalism and of transcendent meaning/purpose. As such, naturalism is a very broad term that would encompass existentialism.

darkness. As from time to time one of these insects draws near it becomes entangled in a sticky thread lowered by the worm, and is eaten. This goes on month after month, the blind worm lying there in the barren stillness waiting to entrap an occasional bit of nourishment that will only sustain it to another hit of nourishment untilUntil what? What great thing awaits all this long and repetitious effort and makes it worthwhile? Really nothing. The larva just transforms itself finally to a tiny winged adult that lacks even mouth parts to feed and lives only a day or two. These adults, as soon as they have mated and laid eggs, are themselves caught in the threads and are devoured by the cannibalist worms, often without having ventured into the day, the only point their existence having now been fulfilled. This has been going on for millions of years, and to no end other than that the same meaningless cycle may continue for another millions of years.[4]

According to the naturalist, the human species is similar in that we continue to reproduce and spread across the face of the earth with nothing accomplished in the end:

> Men do achieve things–they scale their towers and raise their stones to their hilltops–but every such accomplishment fades, providing only an occasion for renewed labors of the same kind.[5]

In claiming that human existence is meaningless, the naturalist is denying that human existence has transcendent meaning, not that it has subjective meaning. Thus even if human existence lacks transcendent meaning, that doesn't mean that human beings can't live well. In fact, some existentialists argue that

abandoning the idea of transcendent meaning is precisely what enables us to attain well-being through the subjective meanings we create.[6]

Study questions

1. True or false: the question of well-being concerns *human* well-being.
2. True or false: water is a good example of something that has instrumental value.
3. True or false: one's best friend is an example of something that has instrumental value.
4. True or false: subjectivist theories of well-being see well-being as analogous to pain in the sense that we cannot be mistaken about it.
5. True or false: objectivist and subjectivist theories of well-being agree that our most fundamental desires

---

4. such, naturalism is a very broad term that would encompass existentialism. 4 Chapter 18 of Richard Taylor's (1970), *Good and Evil*, New York: Macmillan Publishing Co.
5. *Ibid*
6. Albert Camus suggests exactly this in his 1948 novel, *The Plague*.

define what is good for us.

6. True or false: subjectivists can admit that doing what we want is not *always* in our best interest.

7. True or false: objectivists think that doing what we want is *never* in our best interest.

8. True or false: hedonism is a version of the objectivist theory of well-being.

9. True or false: one point in favor of objectivist theories of well-being is that they allows us to deny that the perpetually high drug addict has high well-being.

10. True or false: Plato thought that immoral (unjust) people were really unhappy on the inside, no matter how happy they were on the outside.

11. True or false: one reason for thinking that morality is not necessary for well-being is that this would explain why we resent immoral people whose lives seems to be going perfectly well for them.

12. True or false: an Aristotelian account of well-being would agree that the wealthy, lazy drug addict is living as well a life as could be lived.

13. True or false: the most common way of explaining transcendent meaning is in religious terms.

For deeper thought

1. Consider Davecat, a man who prefers what he calls "synthetic love" (relationships with an inanimate doll) to "organic love" (relationships with another person). Assuming that Davecat most deeply desires synthetic love over organic and that he is happy and fulfilled, do you think that Davecat's well-being is as high as possible? Why or why not?

CHAPTER 9: POST-MODERNISM

WHAT IS POSTMODERNISM?

Paul Jurczak

We are currently living in a historical period called "Postmodern." What we call "Postmodern" is simply what happens after the historical period called "Modern." In the historical development of Western philosophy, we can see various major transitions. What is typically called "modern" philosophy starts with Descartes around the year 1630. Descartes marks a departure from the older Medieval Philosophy that had dominated European thinking. Medieval thought is marked by its adherence to authorities: the Bible and Plato/Aristotle. With the development of the Protestant Reformation (16th century) the reliance on religious authorities was undermined. As the various Protestant churches developed and fought for power with the older Catholic Church, it became unclear which church (if any) might actually have a correct understanding of Christianity. Also with the advances of science, the older Aristotelian model of the world was collapsing. This problem led Descartes and many other European thinkers away from reliance on religious and classical authority. Descartes is "modern" because he refuses to rely on older authorities and, instead, bases his arguments in human reason.

Thus, **Modernism** is the recognition of the limits of older authorities and the reliance, first and foremost, on human reason.As this "modern" world-view develops, it includes the historical era called the "Enlightenment" with its emphasis on the "universal" values of liberal, secular, democratic Europe and North America. The list of great "modern" thinkers usually includes such men as Galileo, John Locke, Immanuel Kant, and Isaac Newton. The "modern" way of thinking culminates in the late 19th century with a great wave of optimism; the Western world believed that their own way of rational-scientific thinking was transforming the world into a paradise of freedom and technological mastery.

That optimism collapsed in the first half of the 20th century. World War I, the Great Depression, and World War II collectively functioned as an on-going crisis. By the end of the First World War (1914-18) France and been economically devastated, the Ottoman and Austro-Hungarian Empires had collapsed, Germany was in

ruins, the Russian Empire had crumbled and some 15 to 20 million people had died in Europe as a result of the war. Then came the Great Depression (1929-1940) which was the worst economic collapse in modern history. It left tens of millions of people without work or income. Then World War Two (1939-1945) concluded with some 60 or 70 million deaths. The rational-scientific methods of the Western world culminated in atomic bombs capable of destroying entire cities. The willingness of the "modern" world to engage in a "rational" and highly technological frenzy of self-destruction was horrifically obvious. By 1945 the world of the "modern" lay in ruins throughout Europe and much of the rest of the world.

The Postmodern world began developing in the ruins of the Modern

Some late modern thinkers had seen cracks in the structure of the modern. Soren Kierkegaard (1813-1855) saw his world as increasingly de-personalized. Friedrich Nietzsche (1844-1900) saw that the modern world had turned most of Europe into a mere "herd" that had lost its independent spirit. Despite these keen early observers, Postmodernism does not begin until after World War II. The "modern" faith in universal values of progress, science, and democracy had left much of the world in ruins.

Another crisis motivated the collapse of the modern; 20th century science was discovering its own limits. **Heisenberg's Uncertainty Principle** was first introduced in 1927. Werner Heisenberg was an early developer of Quantum Physics. He proved that the more precisely the position of some atomic particle is determined, the less precisely its momentum can be known, and vice versa. This inability to know was NOT some lack of human ability; the human sciences were not in some way needing to be improved. Heisenberg saw the world as a place in which some things are, simply, not knowable. The problem of photons is another example of how the world itself is beyond human reason. Photons are effectively *particles* of light when measured one way and effectively *waves* when measured in another manner. So, the "true" identity of light seems to depend on how we observe rather than on some stable fundamental reality. A whole series of discoveries in physics during the 20th century undermined the scientific certainty of the "Modern" world. The most advanced physics of the 20th century was proving that the nature of ultimate reality was itself uncertain. This problem irritated Albert Einstein (1879-1955) who never fully accepted that some things would forever be unknowable. In this sense, Einstein tried to maintain the values of the modern world, but eventually it became evident that human reason has limits.

Although there are many major markers of our "Postmodern" world, in this chapter we will look at four: 1) the rejection of Grand Narratives, and the subsequent re-structuring of the world as 2), "pastiche", "simulacra," and as schizophrenic and 3), the undermining of traditional power relations through Deconstruction and 4) Feminism.

The rejection of Grand Narratives

In an important way, the "Modern" world had valued universal reason as the key to human fulfillment, but after World War II the Western conception of "reason" itself came to be questioned. Postmodern thinking is often associated with a rejection of grand narratives like "progress," "modernity," and "reason." One of the early proponents of Postmodernism was the French philosopher Jean-Francois Lyotard (1924-1998). He claimed that cultures cohere, in part, because people within a specific culture believe in a dominant narrative. For most Christians of the Middle Ages that narrative was told in the Bible. For the people of Ancient Greece, their dominant narrative was told by Homer in *The Iliad* and *The Odyssey*. For the peoples of Europe and North America living at the end of the 19th century their dominant narrative was that of science, democracy, and rational thought. Even the radical 19th century ideals of "Communism" are part of that older (modern) culture. Karl Marx's narrative of the collective workers of the world overthrowing their capitalist masters and re-building the world as the "workers' paradise" had been a dominant narrative for the Soviet Union and Communist China; Communism is now seen as just one of many old-fashioned stories that has proven itself ineffective. **Grand narratives** are understood by Postmodernists as collective myths that never had a reality; grand narratives were attractive and widely believed but they were, at best, collective delusions and, at worst, impositions of power that went unnoticed. Once dominant narratives are shattered, people enter into a period of great uncertainty, groping for meanings and, perhaps, cherishing the "good old days" when they had a single comprehensive narrative that gave their lives meaning. However, the end of traditional meaning also produces the opportunity for people to create new meanings for themselves.

Postmodernism in the world—pastiche & the movies

With Postmodernism, we leave the certainty of a single, integrated, and sense-making narrative, and we enter into a period cut adrift from certainty, plunged into "multiple, incompatible, heterogeneous, fragmented, contradictory and ambivalent" meanings. The loss of a dominate narrative leaves people disconnected from each other, relying on smaller stories of identity like race, social standing, or hobbies that can only bring smaller groups of people together. Frederic Jameson (b. 1934) sees our Postmodern world as one in which cultures are dislocated and language communities are fragmented; each profession is increasingly cut off from others by its own peculiar jargon and private codes of meaning. We are unable to map our world as it breaks into innumerable minor cliques and tribes. A word that is often used to describe this new world is "**pastiche**"—a creative work (like a novel or movie) that openly imitates previous works by other creators. The "pastiche" relies on its readers/viewers sharing the author's cultural knowledge.

The well-known film, *Blade Runner* (1982), takes up and make clear the idea of a "pastiche." In the film, the language on the streets is known as "Cityspeak" and the main character (Deckard) says, "That gibberish he talked was Cityspeak, gutter talk, a mishmash of Japanese, Spanish, German what have you." In this way,

we have the fracturing of language and communities into a pastiche. Another example of this pastiche occurs early in the movie *The Matrix* (1999). The main character (Neo) happens to own a copy of an important Postmodern book of philosophy: *Simulacra and Simulation* by Jean Baudrillard (1981). Neo hides copies of his illicit computer files within this book. Morpheus quotes from the book when he speaks to Neo about the "desert of the real." One doesn't need to know this book in order to appreciate the movie, but for alert viewers the book functions not only as a "pastiche" (a quotation from another domain) but also as a sign of another important postmodern theme in the movie: how we replace reality and meaning with symbols and signs such that reality becomes a simulation of reality. In *The Matrix* Neo sells illicit computer files that function to give people experience; those experiences are merely simulations of "real" experiences. But once those simulations become so real that one cannot distinguish reality from simulation we have no longer a mere "simulation," we have a "simulacrum."

In the movie *Blade Runner*, the main character (Deckard) is employed to hunt down a group of "replicants" who were created to do dangerous work in the outer parts of our solar system, however a group of replicants has escaped and returned to Los Angeles (the city of angles) to confront their maker: Tyrell of the Tyrell Corporation. The replicants are themselves so close to real that it is almost impossible to distinguish them from real humans. They display intelligence, loyalty, anger, mercy and all those other human attributes, even philosophizing about the meaning of their lives. Thus, the replicants are a simulation that are as real as the humans they are meant to simulate; they are "simulacra." At the end of the movie we are left with a disturbing possibility that our main character (Deckard) has fallen in love with a replicant and is, himself, probably a replicant.

"**Simulacra**" (singular form of the plural "simulacrum") is a Latin word that means similarity or likeness. It implies a reproduction of some original object. But in the Postmodern world it takes on a more disturbing meaning. At some point, our reproductions become so much like the originals that it no longer makes sense to distinguish the original from its copy. The simulacrum can then go on to take the place of the original. The Disney theme parks are often referred to by postmodern thinkers as an example of this replacement function. The Italian writer Umberto Eco says of the various Disney parks that "we not only enjoy a perfect imitation, we also enjoy the conviction that imitation has reached its apex and afterwards reality will always be inferior to it" (*Travels in Hyperreality*). Thus, we have the comment in the previous paragraph by Morpheus in *The Matrix* (quoting Jean Baudrillard)—the "desert of the real." Reality becomes insufficiently entertaining or engaging for us; we seek a hyper- reality in which to spend our time. Disney's "Main Street, U.S.A." is Mr. Walt Disney's idealized version of an early 20th century main street in a mid-sized, Midwestern town. But in Disneyland it is better: there is no crime, no liter, no homelessness, no dishonest businessmen or swindlers.

Another example that may help us understand the "simulacrum" is to think about money. Originally banks kept large amounts of gold and silver to ensure the value of our paper money. By the late 20th century those same banks mostly kept computer records that functioned in the place of silver and gold. When a person needed to buy some object, they could go to the bank and obtain paper money: cash. But today, many people use credit/debit cards (or their smart phones) to pay for the things they buy. Paper money and gold are frequently totally ignored; some stores in Europe have tried to only allow electronic transactions. Thus, the computer code that was originally a "simulation" of cash money (which was originally a "simulation" of gold and silver) is, today for many people, more real than cash. Our debit/credit cards are now "real" money; our credit/debit cards are a "simulacrum."

In the movie *Her* (2013), the main character (Theodore Twombly) has a job writing love letters for people who feel incapable of writing such letters themselves. He falls in love with the operating system of his computer (like Siri or Alexa). As a writer of love letters to be used by people he does not know, the main character shows that an absence of personal relation does not eliminate the need for intimacy. Even a love letter written by a stranger and given to another stranger has meaning. This distancing of intimacy is then pushed even further in a love affair between a human male (Theodore) and the operating system of his computer. A set of complex computer codes are able to simulate an intimate relation to such an extent that Theodore's relation with his computer is his intimate relation; for Theodore, his computer's simulation of intimacy is so much like real intimacy that he falls in love with his computer (the "simulacrum").

The Postmodern as schizophrenic

Another major symptom of postmodern existence (the first being Lyotard's skepticism of grand narratives) is Frederic Jameson's conception of the **"schizophrenic" nature of contemporary life**. Jameson borrows from Jacques Lacan (1901-1981) the idea that schizophrenia is a type of language disorder. We rely on language in order to make sense of the ideas of past, present, and future. When people fail to fully incorporate language into their understanding of the world, they become "schizophrenic," no longer inhabiting a world in which past, present, and future are distinct. Rather, as "Postmodern" people, we live within a world in which such distinctions in time are impossible. As "schizophrenic" we are isolated from one another and cut off from the future and the past; there is only and on-going present. Without a constructed sense of time, people have a diminished personal identity because a large part of that sense-of- self is the "project" in which a human life is engaged. Lacking a sense of the future one is unable to motivate oneself into a higher level of accomplishment. The anecdote of the 32-year-old man, having taken some college classes but never completing a degree, working a low-paid part-time job, and living in his parent's basement is a sign that this Postmodern "schizophrenia" may be more than a humorous joke; it may be more real than we want it to be.

The film, *Memento* (2000), has two timelines, one in color and one in black & white. The film alternates the two. Although the black and white time line is historically first and the color timeline of events occurred later, the color events are ordered in reverse. This shatters the notion of a realistic chronology. Most film critics agree that the viewer of the film is supposed to be confused. This temporal confusion is another marker of a "schizophrenic" postmodern world.

Postmodernism in the world—The Novel

The American novelist William S. Burroughs (1914-1997) is usually classified as one of the most influential of the postmodern writers. We can see his works as "schizophrenic." Among the writing techniques he used was the "cut-up" in which previously written texts on paper were cut-up into words and phrases only to be recombined into totally different sentences.

In Burroughs' later works (1981-87) we see a group of 18th century anarchist pirates who attempt to liberate Panama while a late 20th century detective investigates the disappearance of an adolescent boy. The reader is rip-sawn between the two stories full of homoerotic cowboys, Egyptian gods and putrid giant insects. At times characters' gender shifts. Time, space, and identities are fluid; the commonly accepted realities of the past are ignored. The **postmodern novel** is a sustained criticism of the ideas of realism and objective points of view. The understanding of time itself as a linear progression from past, to present, and then into the future is undermined. In the postmodern novel things do happen and characters do act, but there is no causal connection between the things that happen, and there is no stable or temporal reality to the characters.

In some cases, the author of the postmodern novel comments directly on those events and may parody the actions of his/her character. Within the novel, *The French Lieutenant's Woman*, by John Fowles (1969) the author often breaks into his narrative with his own sense of uncertainty. Fowles says, "This story I am telling is all imagination. These characters never existed outside my own mind." Later the author again breaks into his narrative saying, "perhaps I now live in one of the houses I brought into the fiction; perhaps Charles is myself disguised. Perhaps it is only a game."

Jacques Derrida and Deconstruction

One of the most important Postmodern thinkers is Jacques Derrida (1930-2004). His analysis of language and power have been labelled as "**deconstruction.**" His process of analysis consists recognizing that meanings tend to center on a set of symbols.

Western culture tends to see the world as a set of binary opposites with one privileged term in the center and the other term forced into a marginal role. Examples of this kind of thinking can be found in the following sets of terms: male/female, Christian/non- Christian, white/black, reason/emotion. Within "modern" Western culture, the first term in each of these sets is the dominant term and the second term has been forced into a subordinate role. Derrida claims that Western thought at the deepest analysis behaves in exactly this way. The way in which the privileges are set entitle one group of people over the other, and privileged people have historically often worked hard to maintain their privileges: men over women, Christians over non-Christians, white people over non-white people.

Derrida's process of deconstruction strives to set these binary terms into an unsettled and ongoing interaction. He does not want to reverse the structure of domination; deconstruction is a tactical process of decentering which reminds us of the fact of domination while simultaneously working to subvert the hierarchy of terms. Derrida's deconstruction is a radical rejection of foundational kinds of thinking. One advantage of foundational kinds of thinking is the fixed structure of thought; "modern" people prefer to work and live within a society in which the rules and norms are fixed. To upset that fixity can be deeply troubling. But if one happens to be a member of the group of people who are fixed into a subordinate position, then there is an advantage in undermining the fixed system that oppresses one. Derrida himself was a member of two such groups; he was a Jew within a Christian culture and a North African within a Euro-dominate culture. Thus, he saw himself as marginalized by the fixed arrangement of power within his own life.

The history of Western thinking tends to be highly foundationalist: certain ideas are placed centrally and further thought relies on these foundations. One central idea within this system is logic itself. Derrida calls this Western obsession with logical kinds of thinking "**logo-centrism**." Western philosophy since the time of Plato and Aristotle (4th century BCE) has assumed the existence of essences: a form of deep truth that acts as the foundation for further human beliefs. So, Derrida argues that Western philosophy has been a process of

determining and then speaking directly of these deep essences. Words like Idea, Matter, Authority, the World Spirit, and God have functioned (and to an extent still do function) as foundational essences.

Derrida wants to, first of all, demonstrate that none of these terms can exist purely, rather each term only makes sense within a context that includes its opposite. The "ideal" only makes sense in contrast to the "real." The term "matter" only makes sense as the other side of "mind." Secondly, Derrida wants to subvert the priority of the dominant term and set those terms into an ongoing interaction that will not settle into a new relation of domination and submission. When Derrida writes on the opposition between the terms "male" and "female" (which in traditional Western philosophy privileges the male over the female), he does not want to simply reverse the privilege. He realizes that Western thought largely functions as set of binary categories that give meaning to each other; isolated terms cannot have meaning. Even if we use only the word "male" in a sentence and do not mention "female", the term "male" is itself understood as the opposite of "female" just as the term "female" gets its meaning as the opposite of "male." Derrida does not think that these binary relations can be undone. What can be undone, through Derrida's process of deconstruction, is the domination of one of the two terms over the other.

Postmodernism and the criticism of power

This attention to power is part of Postmodern philosophy. Part of the Postmodern criticism of the "Modern" is that "Modern" thinking left the world a terrible legacy of power through sexism, racism, and colonial domination. Within the "modern" mind- set, men were superior to women and are naturally better suited to roles of power; the white races were the more rational-scientific and, therefore, appropriately given power over the non-white races; the nations of Europe and the United States as key beneficiaries of the rational-scientific world view were seen as justified in colonizing the rest of the world. John Locke (1633-1704) is one of the most influential philosophers of "modern" Europe. He is best remembered for inspiring the ideas Thomas Jefferson wrote into the American Declaration of Independence: that all men are created equal. Locke says that "To understand political power right, . . . we must consider, what state all men are naturally in, and is, a state of perfect freedom to order their actions, and dispose of their possessions and persons, as they think fit . . . without asking leave, or depending upon the will of any other man" (*Second Treatise*, Chap. II, section 4).

However, Locke also wrote the constitution for the American colony of Carolina (1669), which ensures that "Every freeman of Carolina shall have absolute power and authority over his negro slaves ..." (article 110). In addition, Locke owned stock in and thus profited directly from the Royal African Company which ran the slave trade for England. Locke saw no conflict between his insistence on the liberty of white Europeans and the enslavement of Black Africans. Since white Europeans (like himself) were by the terms of his own world view a highly rational and scientific society and the people of Africa were not, Locke saw that the Europeans had the

right to enslave and dominate people who lacked the rational and scientific tools that were central to European superiority.

Also, when Locke says "men" he means white, European, property owning males, not the whole of the human race, and certainly not women.

In fact, most European men found it "natural" that they would have power over women and Africans. Locke is not a postmodernist; he is a "modern" thinker helping Europeans move away from older models of power. He is arguing against the aristocratic power of kings, dukes, and barons. Locke's political arguments did help to liberate a class of land-owning European males from the arbitrary power of aristocrats. But it also legitimated the power of white European males in general. No matter how pleasing the words "all men are created equal" may sound to people of the 21st century, Locke (in the 17th century) certainly did not mean to include women or Africans in his conception of freedom. His words marginalized women and non-Europeans.

Postmodern thinkers do NOT want to denigrate the work of Locke or that of Thomas Jefferson; these two men actually worked to enlarge the number of people worthy of being taken seriously as free and able to hold political and economic power. His words helped liberate middle-class, white, European men from aristocratic domination. But neither Locke nor Jefferson worked for the full emancipation of ALL peoples. Rather, by insisting on the legitimacy of middle class, male, white, European, political power, Locke and Jefferson excluded women and non-Europeans from that same kind of power. From a Postmodern perspective, the liberating vision of Locke and Jefferson was not so much "wrong" as it was radically incomplete.

Feminism and the Postmodern

Since Feminism is one of those critiques of power moving in our contemporary world, it seems natural to see the Feminist movement as part of the larger shift in world view that is Postmodernism. One of the Grand Narratives being undermined by Feminism is the seemingly universal distinction between the male and the female and the traditional gender roles that rely on that distinction.

Judith Butler (b. 1956) argues that we should not radically distinguish between "sex" and "gender." The first is often seen as an irreducible biological category and the second is widely recognized to be socially constructed. But material things (like a human body) are understood through the use of language and thus are (at least to some degree) subject to social construction. Even though Butler considers the word "Postmodern" to be too vague to be useful, she argues that the subordination of women has no single cause or solution; there is no master narrative of "woman" that needs to be overcome.

As we've seen, the term "Postmodern" is not easy to define. Nonetheless, there are certain themes or orientations that are commonly considered to be included within that term. Feminism and Postmodernism

have in common the criticism of traditional sources of power, especially the power that subordinates one sex to another.

Masculinity and Femininity have no universal qualities; they are concepts that only take on a reality as they are taught to young people who then take up and begin to live those roles. These roles are enforced in society; people are punished or rewarded in various ways for either failing or succeeding in properly living the role given them by their biological sex. Judith Butler calls these lived gender roles "**performative**." We are like actors assigned either masculine or feminine roles on the basis of biological sex and expected to perform those roles in our public and private lives. But as "performances" there is no deep reality to these roles we play.

Assigning Blame and the Post-Truth World

There have been numerous articles published in the 21st century claiming that Postmodernism is to blame for all our problems: economic stagnation, cultural relativism, the decline of democracy, social fragmentation, weakening of the family, decline in morality, the existence of alternative facts, and the election of Donald Trump. Postmodernism developed as a criticism of power rather than a tool to further empower the already powerful. When a person holding political power claims that his own views of the world are an "alternative fact," and thus just as legitimate as any other fact, he is using a very old tool in order to hold on to power; he is using skepticism not Postmodern deconstruction. Powerful people have always been able to wield skeptical tools against the claims of others. In the early 1600s the Church (both Catholic and Protestant) was skeptical of the claims of Galileo that the Earth was a planet that orbited the Sun. In the late 1900s most scientists were skeptical of Darwin's evolutionary model. The fact that Quantum physics sees the ultimate foundations of the material world as in some part unknowable, does not undermine the science of chemistry's Periodic Table or the fundamentals of mathematics. Postmodern is not post-truth. *What sets the Postmodern world view apart from older forms of skepticism is the radical attempt to open the dialogue to include those who have been systematically excluded: women, people of color, the poor, people who refuse standard gender identities*. However, by opening this discussion, room is made for the already powerful to force their own agendas. When powerful people use skeptical tools to downplay the postmodern forces that question their legitimacy, they are NOT being postmodern. They are working within traditional (modern skeptical) frameworks of power. By trying to blame the postmodernists who work to undermine the traditional power-holders (i.e., male, white, Euro-American, gender normative) our modern power- holders work to maintain their own power and blame the problems they themselves have created on those who lack power; this is well-known in political circles as "blaming the victim."

Why many people resist postmodernism

Let us understand Postmodernism as our current moment in which there is great skepticism of grand

narratives (like progress, Christianity and capitalism), a pastiche of styles in the arts, one in which our simulations are so real that they are often accepted as reality, an on-going attempt to "deconstruct" power, and the of diminishing power for white males. Let us further understand that many people (especially, but not exclusively, privileged, older, white, males) are uncomfortable in this environment and miss the "good-old-days" in which their power was unchallenged and the world in which they lived made sense. We should not be surprised that many of these people actively resist the Postmodern age in which they live. People who, during the modern period would have identified as "normal" and "rightfully powerful," are forced into spaces with people who are NOT just like them. When traditional privilege is questioned and the boundaries between us collapse, **the "other"** forces its presence on us. Men have to deal with women in power; white people have to adapt to having Black people in power; heterosexuals have to cope with homosexuals in their neighborhoods and families. We should not be surprised that many people have problems in coping with this changing reality.

CHAPTER 10: PROPERTY

THE NATURE OF PROPERTY

Paul Jurczak

When we think about property we often think about the "stuff" we own like our computer, our clothes, and maybe a car and a house. We also tend to think that once we own something we can pretty much do whatever we want with it. I can loan my car to my brother, if I want to. I can throw away an old shirt I no longer want. But ownership is actually a complex idea.

Even if I own a car, I can't park it wherever I want. Also, I am required to pay a registration fee to the state every year, and I must buy insurance (in case I crash into someone), and I am required to perform certain repairs even if I don't want to (headlights must function in order to drive at night). Also, my use of my car is limited: I still have to obey the traffic laws.

Additionally, there are complex forms of property like stock investments in large corporations, real estate, and intellectual property. These forms of property are often co-owned with other people to whom one is legally bound. Many people live in apartments which are rented; paying rent for an apartment gives the renter SOME ownership rights but not others. So, ownership of property is not as simple as it might, at first, seem.

Property is a Right, Not a Thing

Property ownership is best understood not as the ownership of things but as a set of rights to the use of things. A person who owns a house has rights to certain uses of the house but not to others. For houses in zoned communities (most houses) the ownership rights are limited. I cannot tear down my house and put up a movie theater without changing the zoning of the physical property upon which the house is situated. The right to change zoning is a collective right held by the community in which the house is located. We can say that the city in which a house is located has an ownership right over the proper use of that house. The city also has an ownership right to extract certain taxes from the owner. The political philosopher C.B. Macpherson says, "As soon as any society, by custom or convention or law, makes a distinction between property and mere physical possession it has in effect defined property as a right. And even primitive societies make this distinction."[1] This distinction between mere possession and property right is further explained by Macpherson:

> ". . . to have a property is to have a right in the sense of an enforceable claim to some use or benefit

---

1. Macpherson, C. B. *Democratic Theory: Essays in Retrieval.* Oxford: Clarendon Press, 1973.

of something, whether it is a right to share in some common resource or an individual right in some particular thing. What distinguishes property from mere momentary possession is that property is a claim that will be enforced by society or the state, by custom or convention or law."[2]

Using Macpherson's language, we can say that the owner of a car has an **"enforceable claim"** to who may use that car. If the owner does not give you permission to use her car, and you use it anyway, she can go to the police who are empowered to take the car from you, with violence if necessary. This is an "enforceable claim," or what we are calling a "right."

Although within a capitalist system we tend to see property as private property (that is, owned by some person or group of persons), we need to understand that there is a long and continuing tradition of **"common" property** even in highly capitalistic societies: streets, public parks, and other common land, like state forests. Although philosophers living in past ages often took for granted certain forms of common property, we in the 21st century are even more highly aware of the need to protect common property like water and air from pollution.

In this chapter, we will examine two major issues: 1) the nature of "property" as either natural or as a social construct, and 2) the possible reasons to limit ownership of property. As for the first issue we find thinkers including John Locke and Adam Smith claiming that property is "natural," and we find David Hume and Thomas Hobbes claiming that property is necessarily a social construct.

Key Question: Is Property a "Natural" Right or a "Social" Right?

Two of the most influential early modern Western philosophers, Thomas Hobbes (1588-1679) and David Hume (1711-1776), argued that there could be no natural right of property ownership, rather the various rights of ownership were created by the state (or some less formal community) and were, thus, a human creation. This human convention we call ownership of property was begun simply because it was useful.

Hobbes and Hume continue an ancient Christian tradition (substantially developed in the work of St. Thomas Aquinas) in stating that the function of law is to protect our rights to survive. To the extent that property rights help us to thrive, those rights are defended in law; when those property rights harm the community, those rights are properly curtailed. Thus, Hume and Hobbes argued that political officials had the right to open granaries and to distribute grain to hungry masses, even if that offended the owners of the granaries. Put simply, if the law stood between a starving person and the food he needed to survive, then the Law no longer had any

---

2. Macpherson, C. B. ed., *Property: Mainstream and Critical Positions*, Toronto: Uni. of Toronto Press, 1978, p.3.

meaningful function and would be superseded by human need. Hume says that the ownership of property "can be done after no other manner, than by a convention entered into by all the members of the society to bestow stability on the possession of those external goods, and leave everyone in the peaceable enjoyment of what he may acquire by his fortune and industry."[3] How exactly this "fortune and industry" made something a person's property is not clear, but Hobbes and Hume insist that ownership serves a public good and that the state needs to establish rules that govern property rights.

There are philosophers who argue that property rights do not derive from human conventions, rather, these other philosophers argue that property rights precede the origins of states and human communities. Arguing in favor of a "natural" origin of property is John Locke (1632-1704). He is probably the most influential thinker on the idea of property. Much of what we might casually say about property rights the is rooted in Locke's philosophy.

Locke argues that although all of god's creation was given to people as our common property, we each uniquely own our bodies and our ability to perform labor. Locke says

> Though the Earth . . . be common to all Men, yet every Man has a Property in his own Person. This no Body has any Right to but himself. The Labour of his Body, and the Work of his Hands, we may say, are properly his. Whatsoever then he removes out of the State that Nature hath provided, and left it in, he hath mixed his Labour with, and joyned to it something that is his own, and thereby makes it his Property.[4]

Locke is saying that all persons have equal right to the goods of nature until that point were some person engages in some work (labor) to obtain that good; in performing that labor the person has joined their personal labor (which is uniquely their own) to some object in nature, and thus that particular object in nature has become the property of that person. We should be able to easily imagine such a case. Imagine a primitive forest in a time before the organization of the world into states, and imagine there were trees in that forest that bore fruits. If some ancient person went into the forest (which exists, according to Locke in a state of nature and thus is owned by no one) and that person were to pick fruits, those fruits become the gatherer's property because that person gathered them. We can also imagine a large lake full of fish. If some ancient person were to put her own labor into catching a fish, that fish becomes her property because the fish in her hands is no longer in the state of nature (where god had originally placed it) but has been caught though her own efforts. So, Locke argues that when any person mixes their own labor into the natural thing, that thing now becomes the property of the person whose labor has been mixed with it.

---

3. Hume, *A Treatise on Human Nature*, 1739, Part II, Sect. II.
4. Locke, John. *Two Treatises of Government*, 1690. (Part II, para. 27) [Spelling and capitalization are from Locke's original text.]

For Locke, ownership of land is somewhat different than the ownership of more moveable things. When gaining a property right to a piece of land, Locke insists that we not simply take from the common store of nature, but that we engage in productive labor. To claim a piece of land from out of the state of nature, our labor must actually increase the amount of goods made available for people. It is not enough for me to harvest the apples from trees in the forest in order for me to claim that I thereby own the forest. According to Locke I must also cultivate that land so that it produces more apples. If by so doing I actually increase the number of apples available for people, then I can call that apple orchard my property because there are now more apples than there had been in the state of nature. So, I can make a section of forest or a coal deposit or a lake my own property if, through my labor, I increase the quantity of goods available. Put simply: a person makes a piece of land their own by increasing the productivity of that land. This way of thinking supported the European colonists in the Americas as they drove out the native Americans and seized their land; since the Europeans could build productive farms, cities, and harbors in what they considered "waste land" (land that the Native American had not made "maximally productive") the Europeans claimed a property right to that land and its improvements. Locke's model says that whoever can maximize the productivity of any given property has the right to do so as long as no one else had already improved that land. European settlers, for the most part, saw the Americas as "waste" land, open to themselves, and everyone else, to seize and improve.

Although Locke's model may seem attractive, it does not seem to have much to say to us in the 21st century. We do not find ourselves in a world full of primitive forests and pristine lakes. Rather, everything I see is already owned by someone. Locke was living in England in the 17th century; even 350 years ago this kind of untamed natural world was not something Locke actually saw. Once most of the world is owned, the abilities of later generations of people to gain ownership rights to the things they want or need has changed radically. We do not live in the world described in Locke's "state of nature."

Locke grounds his theory of property in the Bible. Locke says, "God and his reason commanded him [humans] to subdue the Earth." Thus when "He that in Obedience to this Command of God, subdued, tilled and sowed any part of it, thereby annexed to it something that was his Property, which another had no title to"[5] Further Locke says that god gave land "to the use of the Industrious and Rational."[6] Since the native population of the Americas were not thought to have truly settled and cultivated the lands, it seemed natural to the Europeans (from Locke's point of view) that they had God's sanction to the land of the Americas.

David Hume (1711-1776) claims—much as did Thomas Hobbes (1588-1679) —that there is no natural right to property, rather, all property is grounded in the laws of a just society. Hume says,

---

5. Ibid., § 32
6. Ibid., § 33

A man's property is some object related to him. This relation is not natural, but moral, and founded on justice. Tis very preposterous, therefore, to imagine, that we can have any idea of property, without fully comprehending the nature of justice, and shewing its origin in the artifice and contrivance of man. The origin of justice explains that of property.[7]

Hume brings us back to the idea that the rules that establish property ownership are constructed by a society in order achieve some positive goal(s). If a group of people are to agree that some person actually does have property rights in some object, the conditions whereby that right is properly held must be known and agreed to by that group. Rather than tracing ownership back to some mysterious "state of nature,"

Hume claims that rules of justice must already exist prior to any claim of property ownership. Remember that if we think of property as an "enforceable claim" (and most contemporary philosophy of property does) then there must be a group of people (say police and the courts) that exist in order to enforce one's claims to property.

The Essential Distinction

We see an essential distinction between these two conceptions of property. For Locke, ownership precedes all forms of social connection; property is a **"natural" right**. Laws produced by societies may function to protect natural rights but do not, themselves, produce those rights. The rights to property exist outside of any form of social agreement. From the Lockean perspective, the written laws of property in any given nation can be either moral or immoral: the human laws may either enforce natural law or abridge natural law. A contemporary Lockean might claim that Capitalism merely enforces the natural law of property while Communism abridges those natural rights. However, the claim that property rights are "natural" is not an argument; it is merely a claim.

Hume, on the other hand, sees human society (community) as fundamental. So, for Hume, society has both a temporal and an ontological priority over property. That is to say that there can be no property without first there being some sort of community of persons who can agree upon the rules of property ownership. In this view, there is no "natural" way that property exists; it is merely a human agreement—though, perhaps, a very important one. No set of property rules is natural, since all rules of property are devised and enforced by community. Each community has the right to choose the rules of property for themselves. Therefore, the rules (or laws) of property if freely chosen by a community cannot be, themselves, either moral or immoral. Those property rules are merely the ones chosen and can be changed as the community decides to change them.

---

7. Hume, *A Treatise of Human Nature*, Part II, Sect. II.

Criticism of Locke and Natural Property

Hume is critical of Locke's claim as to the ownership of property through the "mixing" of our labor with some object of nature. There are various kinds of properties people claim to own but have not mixed their labors to it. Hume gives the example of a meadow on which one places one's cattle to graze. In this instance, there is no "mixing" of labor into a natural thing, rather what has happened is that the herder has merely "accessed" a piece of land and not mixed his/her labor into that meadow.

Further, Hume claims that we never actually mix our labor into anything; our labor merely makes some alteration in the natural thing. For example, the fish I catch from some lake is, itself, still a fish, and I have not mixed my labor into the fish itself. I have merely altered the placement of the fish out of the lake and into my hands. The fish itself remains unchanged; it remains a fish.[8]

In support of Hume's position, we should easily be able to think about the kinds of labor we expect human beings to perform that do NOT give any person a property right. If during some particularly cold winter day (say a January blizzard in Michigan) I encountered another human collapsed in the snow, I would have a moral obligation to aid that person. I might use my "labor" to call the police, or I might labor to move that person into a warm shelter. My labor does not give me any property right over that

person I have aided. Justice itself seems to demand that I use my labor to help other people in mortal danger. We can also imagine a cabin in the wilderness of the Upper Peninsula of Michigan. There are people who own such cabins and (because of traditional values) do NOT lock the door to their cabin. The reason is simply that any person lost and in need of shelter has the "right" to make use of that cabin in circumstances where his/her life might be in danger. In this sense of justice, the legal owner of the cabin retains full ownership rights to their property. However, a person whose life might be in danger has certain limited property rights to the cabin as well. *Hume would argue that the cabin owner's right to property is not more important than the lost traveler's right preserve their own life.*

Thinking about property having some natural origin in labor leads to certain problems. First, if while traveling I were to lose a $5 bill, Locke would claim that this specific $5 bill remains mine forever. If you were to find that $5 bill at some future date, you cannot claim it as your own. Abandoning Locke's "natural right" view and taking on a more modern view (property as an enforceable right), it should be clear that since I cannot distinguish one $5 bill from another, I have no "enforceable" claim on the $5 you have found.

Additionally, Locke's picture of natural rights seems unable to explain why people over time so frequently change the rules of ownership. One might think that if ownership were simply "natural," people would figure

---

8. *Treatise of Human Nature*, Part II, Sect. III, FN. 16

this out and then rules of ownership would remain stable over time. However, if we examine almost any set of property rules we find them often changed.

Even though John Locke insists that we all have ownership rights in our own bodies, throughout most of European history women were not allowed to own property or to choose a marriage partner. Several U.S. states changed their property laws in 1835 and for the first time allowed married women to own and manage property in their own name IF their husband were incapacitated. In 1860, the state of New York took leadership in women's rights by passing the New York Married Women's Property Act; this new law allowed married women (rather than their husbands) to control the money they earned. In 1870, the United Kingdom passed the Married Women's Property Act which allowed married women to legally own the money they earned and to inherit property. Prior to these laws, husbands owned not only their own property but also all property that had come to their wives.

We look at these laws simply in order to show support for Hume's position: ownership doesn't seem to have some basis in nature, but, rather, only in human ideas about justice. If we agree with Hume, then we cannot appeal to some original/natural facts about property. We need to think about communities and nations choosing rules that define the terms of property ownership. Those rules would be, in part, the result of some conception of what is justice. Is it justice that allows married women to control the money they earn from their own labor? Is it justice that allows married men to take control of money earned by their wives? As our sense of justice changes, our ideas of property ownership change. When slavery was eliminated in the United States in 1865, it was no longer considered justice for one person to own another as property. During the American Civil War, many people believed and were willing to fight to the death over their sense of their rights to own other people as property. How many contemporary people honestly believe that we may own other people as slaves?

Limitations on Ownership of Property

Human history has been lived in scarcity, and under such circumstances it made sense to claim that if some person owned more, some other person must own less or nothing. It is only since the time of Locke and afterwards that it was thought rational to desire without limits. This desire without limits could only be seen as rational once the industrial revolution had made the elimination of scarcity thinkable. In the history of scarcity, unlimited wealth could never be attained, and, so, as an unattainable goals the desire for unlimited wealth was irrational. However, "What was new from the seventeenth century onwards, was the prevalence of the assumption that unlimited desire was rational and morally acceptable."[9] But humans are complex beings for

9. Macpherson, 1973, p. 18

whom economic development is only part of life. There is a long tradition going backto, at least, Aristotle that "what had to be maximized was each man's ability to realize his essence. What had to be asserted was the equal right of every man to make the best of himself."[10] Included within that unlimited ownership of property is the right to hire other people to perform work in exchange for wages; in hiring others one transfers or appropriates the powers of other persons to one's own agenda. "The unequal properties acquired in market operations became the means by which some men increased their powers by acquiring the powers of others."[11] This kind of transfer increases the ability of one person to attain their goals, but it lessens the ability of the other person to pursue and attain their own highest development. To see Macpherson's point, all one need do is to imagine a small group of very wealthy people who use their powers to create laws that keep unions weak or illegal and wages low. In this a situation, the right to personal development of some has been sacrificed to others.

Our traditional philosophy of property developed in Europe in the 17th and 18th centuries. At this time in European history, philosophers (like John Locke) were engaged in revolutionary thinking; they were arguing against the absolute right of kings. The American legal scholar Morris Cohen explains, "As is natural in all revolts, absolute claims on one side were met with absolute denials on the other. Hence the theory of natural rights of the individual took not only an absolute but a negative form: men have inalienable rights, the state must never interfere with private property, etc."[12]

Ayn Rand (1905-1982)

Although she is not counted among the great philosophers, her published novels and film scripts make her rather well known today. Her thinking on the nature of property is, for the most part, derived from both John Locke and Adam Smith. So, although she is not an original thinker, she brought important philosophical ideas to a wide audience through her works of fiction (mostly *The Fountainhead* and *Atlas Shrugged*). She explores the idea of property in a number of her writings. The following quotation is from her character John Galt in *Atlas Shrugged*, "Just as man can't exist without his body, so no rights can exist without the right to translate one's rights into reality—to think, to work and to keep the results—which means: the right of property."[13] It seems that Rand is echoing the ideas of John Locke: property comes into being through the physical or intellectual work of human beings. Rand's character, John Galt, further says that "The source of property rights is the law of causality. All property and all forms of wealth are produced by man's mind and labor." Here Rand is taking up the argument of Adam Smith. Smith says, "The property which every man has in his own

---

10. Ibid.
11. Ibid., p. 21
12. Macpherson, 1978, p. 167.
13. Rand, Ayn. *Atlas Shrugged*, 1957

labour, as it is the original foundation of all other property, so is the most sacred and inviolable."[14] Like Adam Smith, Ayn Rand claims that property rights are the most fundamental of all rights; without a secure right to hold property no other rights matter; all other rights are founded upon the right to property. Rand says,

> The right to life is the source of all rights—and the right to property is their only implementation. Without property rights, no other rights are possible. Since man has to sustain his life through his own effort, the man who has no right to the product of his effort has no means to sustain his life. The man who produces while others dispose of his product, is a slave.[15]

Rand claims that property rights are the most fundamental of all rights, and without rights to property no other rights are even possible. One might wonder if rights to religious freedom concern Ayn Rand. They do not. One might imagine that a person who was enslaved (and thus had no rights to any property) might still find value in their "right" to practice a religion of their choice or to live with their family. There is good evidence that enslaved people did value these rights. In contrast, Rand was a committed atheist who had no regard for religious practices of any kind; she left her family in Russia in 1926 and never returned. She says little about religion in her works of fiction. Her most thorough look into religion was an interview she did for *Playboy* magazine (March, 1964). When asked if she thought that religion had done any positive service to humanity, she responded, "Qua religion, no—in the sense of blind belief, belief unsupported by, or contrary to, the facts of reality and the conclusions of reason. Faith, as such, is extremely detrimental to human life: it is the negation of reason." Unlike Rand, many people see that property rights are only one aspect within a larger collection of human rights. Among contemporary philosophers, property rights are usually regarded as important, even central, but property rights are not understood as the sole foundation upon which all other rights are grounded.

Even very conservative thinkers have argued in favor of limiting, at least, some forms of property. For intellectual property (like copyrights and patents) Ayn Rand makes clear that this kind of property right can only have a limited lifespan: the life of the creator of that intellectual property. When the creator dies, the right to full use of that creation spreads to all living people; it does not make sense for the dead to control what the living may do with ideas. Rand makes clear that intellectual property rights cannot be granted in perpetuity. "If it were held in perpetuity, it would lead to the opposite of the very principle on which it is based: it would lead, not to the earned reward of achievement, but to the unearned support of parasitism."[16] Rand sees a use (and most other people do as well) in granting certain limited rights to the creators of ideas and works of fiction. By providing those creative persons the sole right to their creations, society rewards, and thus encourages, creative

---

14. Smith, Adam. *An Inquiry into the Nature and Causes of The Wealth of Nations*, 1776. Bk. 1, Chap. 10, pt. 1.

15. *Ayn Rand Reader*, 1999, p. 380.

16. Rand, Ayn. *Capitalism: The Unknown Ideal*, 1966, p. 131.

work. But allowing copyright and patent rights to live on past the lives of these creators is to grant a right that is unearned. To pass copyright and patent rights across generations is to grant a right to function as a parasite—to be rewarded for work one has done nothing to earn.

Rand is usually considered a politically "conservative" thinker, one who defends the property rights of the wealthy. But, specifically, as to the property we call **copyright**, her argument condemns some of the largest and most powerful copyright holders, including the Disney Corporation. The Copyright Act of 1976 allowed for copyright to last for the entire life of the author plus 50 years, or 75 years for corporate copyrights. However, in 1998 the U.S. Congress passed the Copyright Term Extension Act (CTEA). This act further extended copyright to the life of the author plus 70 years, and the legal copyright of corporate products to 95 years after publication. This act (CTEA) was widely ridiculed as the "Mickey Mouse Protection Act." The Disney Corporation encouraged Congress to pass the law; CTEA allows Disney complete control of all images of Mickey Mouse until the year 2024. Had CTEA not been enacted, Mickey Mouse would have entered public domain and thus been released to the world for free and unlicensed duplication in 2003, thus reducing profits of the Disney Corporation.

This law allows the Disney Corporation to charge substantial fees on all images of Mickey Mouse decades after the creators have died, thus permitting the Disney Corporation to function, in Ayn Rand's words, as a "parasite" by continuing to extract money it has not earned from the character of Mickey. Many conservative economists agreed with Rand and opposed the extension of copyright, these economists included Kenneth Arrow and Milton Friedman.

Rand is of interest in part because of her influence on later thinkers of importance. Even though her influence on academic philosophy has been very minor, she cultivated friendships with people who later did become politically powerful, including Alan Greenspan who served as Chairman of the Federal Reserve bank from 1987-2006.

Rand is often associated with modern **Libertarian** thought; she believed, for instance, in a woman's unlimited right to abortion, and she opposed the Vietnam War. On the other hand, many understand her work as aligned with mainstream conservative thought; she condemned homosexuality and insisted that Europeans (in the 16th through 19th centuries) had the right to take the land from American Indians. She is a complicated thinker who was most successful in providing entertaining works of fiction in which she developed the ideas of others, especially the ideas of Adam Smith and John Locke.

A Christian Point of View

A much older philosopher working in the 13th century explains why property rights are always limited. This philosopher is St. Thomas Aquinas, and his writings are foundational for much of Christian morality.

Although Aquinas lived much of his life in a monastery, he was not cut off from human suffering. He tries to understand (from a Christian point of view firmly grounded in the Bible) what kinds of moral obligations people have to each other. Aquinas has an interesting argument on what prosperous people owe to people in dire need: hunger, thirst, cold and disease. He says " In cases of need all things are common property, so that there would seem to be no sin in taking another's property, for need has made them common."[17] In his argument, Aquinas follows the logic of his Christian belief even when it leads to an answer that some people (especially Adam Smith and Ayn Rand) might find troubling. From his Christian point of view, Aquinas teaches that all useful things have been put on this planet by God to supply the needs of human beings. For example—water in rivers and lakes is present to take care of our thirst. Plants and animals exist to nourish our bodies. Trees grow to supply wood for houses and other human needs. There is such a thing as privately-owned property for Aquinas, and under normal circumstances we have no right to take the property of other people. However, when someone is truly in NEED, property that can address those needs becomes "common property," and private property ceases to exist. Common property is like air to which we all have a common right; no one may claim exclusive ownership over it.

According to Aquinas the person in need should ask the owner of abundant property for help, but if the owner is not present or if the owner denies access to what a person truly needs, the person in need may take and use the needed property. In such circumstances, no sin occurs. What may surprise us is that any person with abundant property who refuses to aid a person in need has sinned by hoarding the property that God has placed onto the Earth to serve all human needs, not just the needs of a few.

Taking the property one needs is no more a sin for Aquinas than is drinking water from a lake; all things are common property when someone is truly in need and so all things are available to whomever needs to use them.

Sir Anthony Kenny (a contemporary philosopher) summarizes the philosophy of Aquinas on wealth and property.

> First of all, it is sinful to accumulate more property than one needs to support oneself, relatively to one's condition in life and the number of dependents one has. Secondly, if one has money to spare one has a duty—as a matter of natural justice, and not of benevolence—to give alms to those in need. Thirdly, if you fail to relieve the poor, then they may, in urgent need, legitimately take your property without your leave.[18]

---

17. St. Thomas Aquinas. *Summa Theologica*, Book II, Part II, Quest. 66, Art. 7.

18. Kenny, Anthony. *Medieval Philosophy, A New History of Western Philosophy*, Vol. 2, Oxford: Oxford University Press, 2005, p. 270.

John Rawls and the Limits on Property

John Rawls (1921-2002) argues that a society built upon unfair principles for the distribution of property is not a society we would choose to be part of and is, thus, an immoral society.[19] Thus, he agrees with Hume, Hobbes and Aquinas that ownership of property is not a primary right; rather, it is based in a society's agreement on the notion of justice. What one may rightfully claim as one's property is grounded in a society's understanding of justice. Rawls develops his ideas on property from within a more general discussion of justice. Rawls acknowledges that within any society, there will almost certainly be people who are not well off. But a just society will not permit people to be systematically discriminated against or impoverished. Thus, Rawls insists that IF we were given the opportunity to choose which society we were to be born into, we would choose one that abided by what Rawls calls "**The Equality Principle**"— All offices and positions are to be effectively open to all people regardless of social background, ethnicity, sex, or other possible distinguishers. If I did not know whether I was going to be born male or female, if I didn't know whether I was to be born into one ethnic group or another, I would choose to be born into a society that operated in accordance with "The Equality Principle." In such a society, all jobs and offices are *effectively* open to all adults within that society.

We recognize that not everyone can be a brain surgeon or a physics professor. Our communities also need people whose careers are more modest and who will likely be paid less. One reason we pay a window washer less than a brain surgeon is that a large percentage of our adult population is capable of washing windows, while a relatively small number of people have the capacity to become a brain surgeon. Also, one must train longer for some jobs than others. Thus, we see very plainly, that Rawls' "Equality Principle" will allow some people to be paid more than others; equality of opportunity seems to lead inevitably to some level of economic inequality. So why would we tolerate such a difference? Is a society that pays brain surgeons $250,000 per year and window washers $25,000 per year a just one? Would you willingly enter into such a society IF you did not know the kinds of skills and abilities you might have? Rawls answers that we would willingly enter that society IF we had a further guarantee. Rawls calls this the "**Difference Principle**"—we accept unequal distribution of offices and positions IF those differences function to benefit the least well off.

In this instance, Rawls argues that we would willingly enter into a society in which his "Equality Principle" and his "Difference Principle" (and also a principle of equal basic liberties like freedom of religion, assembly, and speech) operated EVEN if we did not know what position we might hold within that society. We would agree to enter that society because (even if I became the window washer) having highly qualified brain surgeons in my society benefits me (the window washer) when I have access to the brain surgeon should I need one. Rawls argues that even people who are poorly paid are treated fairly IF all offices and positions of a society are effectively open to all AND IF the poor have some real advantage in having well paid people with useful

---

19. Rawls, John. *A Theory of Justice*, 1971.

skills in their society. Exactly how any society might choose to enact Rawls' "Difference Principle" can vary. Some might choose to more heavily tax the wealthy so that the "least well off" have some benefits they might not otherwise possess. Also, a just society (according to Rawls) might strongly encourage (or perhaps force) people who possess highly paid skills to share those skills with the people who are the "least well off." Either way, we should see that within Rawls' model, the "property" of the well-paid members of a community can be compromised for the purpose of a more just society.

A second look at Locke

Locke (and by extension those philosophies that develop his tradition) has been criticized by more modern thinkers for failing to understand some fundamental aspects of our humanity and for fueling (among other things) the current financial situation in which the distribution of wealth is extraordinarily unequal. C. B. Macpherson (1911-1987) is, perhaps, Locke's most perceptive critic. Remember that Locke claims that each of us has unique ownership rights over ourselves. Locke says my labor, my mind, my ideas are fully my own. "Though the Earth . . . be common to all Men, yet every Man has a Property in his own Person. This no Body has any Right to but himself. The

Labour of his Body, and the Work of his Hands, we may say, are properly his."[20] Macpherson calls Locke's position "**possessive individualism**." Locke fails to see that we exist within complex families and communities which have legitimate claims on us. Locke uses a totalizing metaphor of property that denigrates our humanity in seeing that individual persons relate even to themselves as "property." Within Locke's concept of "self-ownership" individual people view themselves as an aspect of property rather than as complex moral/religious wholes or as part of an encompassing social community. In fact, Locke implies that our primary manner of relating to other people is economic—as nothing more than a set of "relations of exchange." But is that the best understanding of Locke?

Locke and others have assumed that each person is entitled to the full product of their own labor. But such an assumption only engenders more questions. It should be obvious that from within any moderately complex society no one's labor is ever fully their own. We should ask "How shall we determine what part of the value of a table should belong to the carpenter, to the lumberman, to the transport worker, to the policeman who guarded the peace while the work was being done, and to the indefinitely large numbers of others whose cooperation was necessary?"[21] Though we may agree that in some sense we do "own" ourselves and our labor, does that ownership mean that no one can legitimately make any claim on my property or on my labor? Isn't

---

20. *Two Treatises of Government*, II, para. 27
21. Macpherson, 1978, p. 163.

virtually all labor performed within highly social contexts and within series of interconnected processes? In one of the more entertaining explanations of Adam Smith's *The Wealth of Nations*, he states how a single workman, alone, might "perhaps, with utmost industry, make one pin a day, and certainly could not make twenty."[22] However, with "modern" (1776) industry a small shop could, "make among them about twelve pounds of pins in a day. There are in a pound upwards of four thousand pins of middling size. Those ten persons, therefore, could make among them upwards of forty-eight thousand pins in a day."[23] How much more interconnected is a 21st century manufacturing plant? How do we figure the benefits we derive from each step of a complex process of manufacturing? It would seem that there is very little that is "natural" in determining the exact extent to which each person in a complex manufacturing process has contributed their "labor."

One problem in understanding John Locke's writings on property is that people commonly do not read enough of Locke! In 1690 Locke published *Two Treatises of Government*. The "second" treatise is still widely read, whereas the "first" is not. The original title of Locke's Two Treatises is overly long, as was common in his time, but explains why the first treatise is rarely studied any more. The original title is: *Two Treatises of Government: In the Former, The False Principles and Foundation of Sir Robert Filmer, and His Followers, are Detected and Overthrown. The Later is an Essay*

Concerning the True Original, Extent, and End of Civil-Government. Locke's purpose in the first essay is to argue against an older political philosopher Robert Filmer (1588- 1653). Filmer had published a defense of the "**divine right of kings**" to absolute authority over their people. Locke, in the first essay, argued against Filmer's position. Today, we no longer seem to need much of an argument to convince us that kings are neither appointed by God, nor do kings have absolute rights to do whatever they please. Thus, few students read the first treatise. However, when we do read that first treatise we find Locke making some claims that show him to be less of an "absolutist" with regard to property rights. We find that his argument is very much the same as the one we read from St. Thomas Aquinas.

> But we know that God has not left one Man so to the Mercy of another, that he may starve him if he please: God the Lord and Father of all, has given no one of his Children such a Property, in his peculiar Portion of the things of this World, but that he has given his needy Brother a Right to the Surplusage of his Goods; so that it cannot justly be denied him, when his pressing Wants call for it. And therefore no Man could ever have a just Power over the Life of another, by Right of property in Land or Possessions; since 'twould always be a Sin in any Man of Estate, to let his brother perish for want of affording him Relief out of his plenty. As Justice gives every Man a Title to the product of his honest Industry, and the fair acquisitions of his Ancestors descended to him; so Charity gives every Man a Title to so much of

---

22. Bk. I, Ch. I
23. Ibid.

another's Plenty, as will keep him from extreme want, where he has no other means to subsist otherwise; and a Man can no more justly make use of another's necessity, to force him to become his Vassal, by withholding that Relief, God requires him to afford to the wants of his Brother, than he that has more strength can seize upon a weaker, master his to his Obedience and with a Dagger to his Throat offer his Death or Slavery.[24]

So Locke, who some have argued claims a right to property that is absolute, actually makes no such claim. To the laws of justice, Locke has mated the laws of charity, which equally bind all people. Locke agrees with Aquinas that prosperous people have obligations to the less fortunate. In more modern terms, Locke claims that we do have property rights, but along with those rights come *responsibilities* to assist the least well off.

Locke is aware that many people will sell their labor to others. He doesn't seem to think that such a situation violates his "labor theory" of property. In the very beginning of the property chapter of the *Second Treatise* (Chapter V) when Locke is establishing the connection between labor and property rights he says, "Thus the grass my horse has bit; the turfs my servant has cut; and the ore I have digged in any place where I have a right to them in common with others, become my property." It might seem strange that in attempting to show how labor creates property that Locke's chosen example is of a servant's labor belonging to the employer. Locke seems to have treated it as entirely natural and understandable. If each man has a property in his own person, he has the right to sell the use of that property if he so wishes. However, Locke limits the extent to which any man may sell his labor to others; that limit is reason.

What might be even more disturbing is Locke's support for and profiting from the institution of slavery. Locke owned stock in the Royal African Company which was begun in 1660. This company was originally granted the right to exploit the gold fields of West Africa, but in 1663 it began to deal in slaves. In the 1680s the Royal African Company was transporting some 5,000 slaves per year to the Americas. Locke was also the author of the document upon which the American colony of Carolina was founded. This document is called the *Fundamental Constitutions of Carolina* (1669). Though it has many disturbing claims, most critics of John Locke point to article 110: "Every freeman of Carolina shall have absolute power and authority over his negro slave, of what opinion or religion soever." Locke not only profited directly from the trade in slaves but he formulated laws that enshrined the rights of slave owners to absolute authority over their slaves. Locke's distinction of who is free (and therefore able to hold property) is based in "reason."

When writing about the relations between husbands and wives Locke states that since someone must rule, "it naturally falls to man's share, as the abler and stronger."[25] Although Locke does not here use the word, he sees

---

24. *First Treatise*, Chap. IV, § 42.

25. *Second Treatise*, § 82.

men as abler in the use of "reason" than women. He makes the same defense of parents over children. However, when it comes to slaves, Locke says

> But there is another sort of Servants, which by peculiar Name we call Slaves, who being Captives taken in a just War, are by the Rights of Nature subjected to the Absolute Dominion and Arbitrary Power of their Masters. These Men having, as I say, forfeited their Lives, and with it their Liberties, and lost their Estates; and being in the State of Slavery, not capable of any Property, cannot in that state be considered as any part of Civil Society; the chief end whereof is the preservation of Property."[26]

By placing slaves totally outside of "civil society" Locke has allowed harsh treatment and substantial financial reward for the masters. Might we wonder how many slaves actually were "captives taken in a just war." And even if this were the case, how can Locke justify his placement of them radically outside of any civil society simply because they are not allowed to own property?

Ultimately, Locke grounds property rights in human reason. Those persons regarded by Locke as having developed their reason were free, while those persons deficient in reason were not. The most central freedom accorded those people in possession of reason was the freedom to own property. "The Freedom then of Man and Liberty of acting according to his own Will, is grounded on his having Reason, which is able to instruct him that Law he is to govern himself by, and make him know how far is left to the freedom of his own will."[27]

What Do We Owe to One Another?

Even in our own society it is rare to hear people proclaim that it is unjust to moderately tax a wealthy old bachelor to help pay for public schools, roads, or local fire departments. We know that paying moderate taxes is fair because of the interdependent nature of our communities. No one can justly claim that their wealth is entirely the result of their own unaided effort.[28] Within a Judeo-Christian ethic of property ownership, we should always remember, and St. Thomas Aquinas reminds us, that the first right to property is based in need not in labor. Even Locke manages to justify private ownership of land only because a larger social need is served; Locke says that I can justify my claim over a parcel of fruit trees in a forest only because with my ownership and hard work more fruit is available to the other people in the neighborhood. If I merely gather the fruit and do nothing to increase the productivity of the piece of forest, I can claim only the fruit I labored to gather; I cannot also claim an ownership right to the forest itself. According to Locke, any ownership claim to a piece of land can be only justified through hard work that makes the land more productive and thus able to serve more

---

26. Ibid., § 85.
27. Ibid., § 63.
28. Macpherson, 1978, p. 163

human needs. Thus, Locke is himself immersed in a Christian ideal grounding ownership in need rather than in labor.

The error some people make in their understanding of Locke stems from an incomplete reading of Locke and in thus reducing human beings to property. Locke is often interpreted as having a reductive understanding of our humanity. Locke's writings imply that our first relation to ourselves is as "property": we own ourselves. As we saw above, C.B. Macpherson labelled Locke's idea of self-ownership as "possessive individualism." For Locke we are essentially what we own, and our society is nothing but the sum of the "relations of exchange between proprietors." Macpherson's essential criticism of Locke (and of all theories of property based in Locke) is this impoverished concept of self-ownership in which every individual sees him/herself as a "property" rather than as a moral whole or a part of a larger social whole. However, reading Locke's First Treatise of Government we see that, like Aquinas, Locke himself held a complex understanding of human nature grounded in Christian theology.

We should not dismiss **Locke's "labor theory" of property**. It does contain an important idea. Labor is encouraged if the people who labor are rewarded for their work. And the distribution of property in accordance with the amount of one's labor does function to encourage hard work. But since property is legitimately only one human interest— alongside of justice, intellectual and physical development, freedom of conscience and of religion—the right to property cannot be pursued if it leads to harming the interests of others. If we are to be thoughtful about the issue of property, we should think neither of prohibiting private property altogether nor of allowing the pursuit of property in a fully unregulated way. We should encourage labor and the acquisition of property within the boundaries of the concept of the "common good."[29]

Questions for deeper thought

1. Why does it make sense to think about property as a "right" rather than as a "thing"?
2. Explain Locke's "labor theory" of property. Then explain the strengths and weaknesses of Locke's theory.
3. St. Thomas Aquinas argues that in some cases people taking the property of others is not sinful. Explain Aquinas' argument.
4. Is property a "natural" right or is it a man-made right?
5. In what way does Ayn Rand want to limit ownership of property?
6. What problems (both philosophical and practical) arise if we think of property as the original right upon which all other rights are grounded?
7. How does Locke ground property rights in "reason"?

---

29. Ibid., p. 167

8. What are some of the ramifications of Locke's understanding of who has reason?

9. John Rawls' claims that unequal distributions of property can be done in a just manner, however if such a society is to be just then it must also include both his "Equality Principle" and his "Difference Principle". Explain briefly these two principles and why Rawls claims they are necessary for a just society.

CHAPTER 11: CAPITALISM, COMMUNISM, & BERNIE SANDERS

CAPITALISM, COMMUNISM & BERNIE SANDERS

Paul Jurczak

Who deserves what?

The three major modern forms of economic production are (in a very real way) all systems by which societies decides "who deserves what." When people claim that an employer has not paid them fairly, when people claim that the system is rigged against them, when people complain that the CEO of a major corporation was paid $10 million for a single year's work, when people see prosperous international corporations paying no taxes, people are complaining about the economic system within which they work and live.

In the modern world there are only three major economic systems: Capitalism, Communism and Socialism. The origins of capitalism are most usually traced back to the 17th century Dutch state. Modern Communism usually traces its history back to Karl Marx in the mid 18th century and began as a rejection of Capitalism. Socialism has a double origin; many trace its modern beginnings to the French Revolution (1789) and then again as a hybrid between Capitalism and Communism in the later 19th century.

We can formulate a very basic statement about "who deserves what" for each of these economic systems:

   a. **Capitalism**—To each according to what s/he and the capital and land s/he owns produces.
   b. **Communism**—To each according to her/his need.
   c. **Socialism**—To each according to some mixed formula of production and need.

Milton Friedman (1912- 2006) is one of the staunchest supporters of Capitalism. Milton Friedman describes Capitalism as the principle "To each according to what he and the capital and land he owns produces."[1] Thus Capitalism is an economic system based in the private ownership of the "means of production" (i.e., capital) and use of capital for profit. Within Capitalism, each person is typically thought to own his/her own labor as well as whatever "capital" (money, land, shares of corporate stocks and bonds, and various other property rights) s/he may have accumulated. What we often call a **"free market" economy** is a form of Capitalism that

---

1. *Capitalism and Freedom* (Chicago, 1962), pp. 161-2.

is only very little regulated by the government and thus allows individuals and groups of individuals to decide on what to produce so as to maximize their own profits.

Karl Marx (1818-1883) is usually credited with the development of Communism. Communism teaches that under Capitalism a great social and political conflict arises between the class of owners and class of workers. The Capitalist class use their amassed **capital** (money, factories, land) to oppress the workers, who (lacking capital) must either labor or die. Because the **working class** is always larger than the **capitalist class**, the needs of the bulk of the population are sacrificed to further increase the wealth of the capitalists. According to Communism this conflict can only be ended with the elimination of that class distinction; people work but not within a system of social classes. This is to be accomplished by the elimination of individual ownership of capital; all means of production (capital) are to be owned collectively by the state or some more limited group of workers. Thus, in Communism, capital is collectively owned by the citizens of the state rather than the property of a relatively small group of wealthy persons.

Socialism is an economic system characterized by some mix of private and social ownership of the means of production and workers' self-management. Socialism has many varieties, but most modern forms include some limited participation in capitalist markets. That is, Socialism is most usually a hybrid system that combines social and personal ownership of capital.

When people complain about the "evils" of Socialism they often overlook the fact all wealthy countries already have a hybrid system within which private capital and social ownership combine. All we have to do is to ask a few questions to see this mix. Who owns the local fire department? Who owns the roads? These are, of course, owned socially; most cities "own" their own fire department for the good of all its citizens. The roads are "owned" by the city, or the county, or the state, or (for interstate highways) the federal government. They are operated for the benefit of all citizens.

Even in Capitalism Some People Are Exempted from the Terms of Capitalism

Every society has groups of people who will be supported (to some extent) NOT by their own work or from the proceeds of their capital. Children in the advanced affluent societies are not expected to earn their own living. People with severe mental and/or physical disabilities, the elderly, severely wounded military veterans, these are all people who are either incapable of supporting themselves or are not expected to support themselves. This being the case, we should see that Capitalism does NOT work for such people, and this is true simply because these people cannot generate a salary sufficient to pay for their own needs—typically food, shelter, and basic medical care. Remember the staunchest defenders of Capitalism have defined Capitalism as an economic system in which each person is paid "according to what he and the capital and land he owns produces." So, in instances where people cannot perform labor (because of severe mental or physical disability) AND these same

people lack sufficient capital (investment income) to be able to support themselves, then the rest of the society does NOT simply let these people die. Rather, various "welfare" programs are instituted, and these programs are most frequently paid for through taxes. The point being made is, simply, that even in fully Capitalist systems it is NOT expected that all members of that society will function within the terms of Capitalism; rather some set of people are exempted from the terms of Capitalism because some people would simply perish within a purely Capitalist system.

Is Capitalism Required for Human Freedom?

Milton Friedman defends Capitalism on moral grounds as the ONLY economic system that can maximize human freedom, and the maximization of human freedom is (for Milton Friedman and most defenders of Capitalism) the highest and only argument necessary to defend Capitalism. Friedman is very concerned about concentrated power. He sees concentrated power as a threat to freedom. This is not an unusual argument. The Constitution of the United States of America itself is grounded upon exactly this threat of concentrated power. A system of "checks and balances" by which the Executive, Legislative, and Judicial branches of the American government maintain their independent powers is an essential aspect of the Constitution. Thomas Jefferson wrote about the need to divide these powers in his "Notes on the State of Virginia" (Query XIII) in 1794, "The concentrating these [powers] in the same hands is precisely the definition of despotic government." In a similar fashion, Friedman is concerned about the combination of economic and political power; an essential role of Capitalism is, according to Friedman, to prevent such a combination of powers. Friedman says, "the kind of economic organization that provides economic freedom directly, namely, competitive capitalism, also promotes political freedom because it separates economic power from political power and in this way enables the one to offset the other."[2] Friedman worries that combining economic and political power in the same hands would become highly coercive and, thus, harmful to freedom. Friedman states that by separating economic and political power, Capitalism "enables economic strength to be a check to political power rather than a reinforcement." [3] Does this actually happen?

Friedman claims that the only economic system that can operate without coercion is Capitalism. His argument begins with an illustration of a rather strange society. In a simple market system each household controls a set of resources (money, food, tools, clothing & other items useful for human life). And each household chooses to engage in trade with neighboring households OR chooses not to. Each household is free to choose because it produces everything it needs. Friedman states that

---

2. Ibid., p. 9
3. Ibid., p. 15

. . . since the household always has the alternative of producing directly for itself, it need not enter into any exchange unless it benefits from it. Hence no exchange will take place unless both parties do benefit from it. Co-operation is thereby achieved without coercion.[4]

This is a very odd community; one in which each household produced sufficient food, housing, clothing, medicine, and computers such that they did not need to engage in trade with their neighbors. Nonetheless, Friedman proposes exactly this "household without need" as his basic economic unit. Unfortunately, his vision of a set of households each fully independent of all others cannot be found in any historical period anywhere in the world. Friedman's vision seems to operate as did the "State of Nature" for both John Locke and Thomas Hobbes: a *hypothetical* original position (or reference point) during which all people were (theoretically) fully free and independent. Friedman then attempts develop an economic system (Capitalism) that retains that original state of near total freedom.

For Friedman people retain their freedom so long as they are not subject to coercion. If we live and work within an economic system that forces our actions, we are not free. Friedman admits that some limited coercion is necessary for any complex society. Should we decide that everyone must pay some limited taxes in order to fund the basic functions of the state (road construction, law enforcement, fire departments), then those who refuse to pay would have to be coerced. Friedman states on many occasions that Capitalism is NOT **anarchism**; any society will have, at least, some few essential rules and these rules will be enforced by even a very limited government.

Friedman's claim that only Capitalism allows for economic exchange without coercion is grounded in the thought of John Locke (1632-1704). According to Locke the purpose of government is very limited. "The great and chief end therefore, of Men uniting into Commonwealths, and putting themselves under Government, is the Preservation of their Property. To which in the State of Nature there are many things wanting."[5] Locke presumes (as does Friedman) an original state of human beings living without law or government of any kind; Locke calls this a "**state of nature;**" Friedman postulates a "household without needs." The problem of living in a "state of nature" is that although people are absolutely free to do as they please they are also living very insecure lives. They may be attacked at any moment; their property may be destroyed or stolen. This absence of security is what motivates people to freely join together into some sort of "government." This union, or as Locke calls it "commonwealth," has the great advantage of making our property more secure, but in order to gain that security, we must surrender some degree of our freedom. Locke claims, and it is easy to agree with him, that no person willingly surrenders all of their freedom and willingly becomes a slave. For Locke some of freedom in order to attain some increased security is an unavoidable choice;

---

4. Ibid., p. 13.
5. Locke, *Second Treatise*, Chap. IX, § 24.

living in the "state of nature"—that is without government—is simply too dangerous. Locke thinks that to give more power to government is to diminish the freedom of the people who formed the government. Friedman agrees with Locke that the only government a person would willingly enter into is one with minimal powers. Friedman and Locke assume that by limiting the powers of government we maximize the natural freedoms and powers of individuals; every increase in government power must decrease the freedom of those people living within government.

We should remember that Locke, in 17th century England, was living through a major political/economic transition: new democratic/capitalist ideas were growing stronger as older medieval/feudal ideas were weakening. Locke's political and economic thought are a direct assault on the absolutist powers held by medieval kings. From Locke's place in history, he sees that government is both the single most pressing threat to human freedom and a necessary safeguard for our property. His solution is minimal government. Friedman's arguments grow directly out of this Lockean conception of the dichotomy inherent in all government, and his solution to this problem is the same as Locke's: minimal government. Since both socialism and communism seem to demand a powerful government, Friedman sees these two economic structures as incompatible with human freedom. This is not the place for a full discussion of the nature of government, but later thinker (including John Dewey and John Maynard Keynes) have put forward cogent arguments that this dichotomy between government power and individual freedom is false.

The Nature of Coercion

Moving beyond his initial model of the "household without need, Friedman goes on to develop an economic model suitable for a complex, modern society. He claims that within a complex modern world, economic exchanges are voluntary (lacking coercion) if two conditions hold:

> (a) that the enterprises are private, so that the ultimate contracting parties are individuals and (b) that individuals are effectively free to enter or not to enter into any particular exchange, so that every transaction is strictly voluntary.[6]

First of all, why need the enterprises be private? If I wish to enter one of Michigan's beautiful state parks, I'm required to pay a fee. If I wished to go to a private park (like Disneyland), I would also pay an entry fee. Why is my entry into Disneyland not coercive while my entry into a Michigan state park coercive? Friedman is concerned that there be numerous market choices; if some super powerful government owned all the parks I would have no choice but to contract with the government if I wished to visit a park. In the realm of parks, I can choose one of the numerous county parks rather than a state park, or I may choose a city park, or even

---

6. *Capitalism and Freedom* (Chicago, 1962), p. 14.

a national park. No single government owns them all. But still we should say that none of those parks meet Friedman's requirement for ownership by "individuals," since they are owned by various levels of government. But it is unclear how the fact of community (or government) ownership of the parks makes me less free. However, if all the parks were owned by some monolithic, evil, government, that government could (in theory) ban me from all its parks. Friedman seems to assume that under Communism or Socialism some monolithic government would own everything. This might have been true under the most autocratic period of Soviet Stalinism, or Mao Zedong's China, but it simply isn't true of any modern, democratic, nation, even the ones some people call "socialist," like Sweden and Denmark.

Friedman's second requirement is that for any person to be free they must be able to choose NOT to enter into any "particular exchange." One of the key properties of any Capitalist society is the existence of a group of laborers who lack sufficient capital to support themselves and thus have no choice but to enter into the market to sell their labor. These workers may either work or starve; such is the coercive element of Capitalism. Notice that Friedman tries to avoid this conclusion by limiting coercion to situations in which a person is "free" to NOT engage in some *particular* exchange.

Since the worker may have a choice of who to work for, Friedman sees the worker as "free" because the worker seeking employment may choose NOT to work for one particular employer because there is always another employer hiring employees. From Friedman's perspective, a particular worker is NOT coerced into selling his/her labor as long as that worker may choose to sell his/her labor to either Wal-Mart or MacDonald's. What Friedman fails to consider is not only that under circumstances of high unemployment there may only be a single employer (or no employer at all), but (more importantly) the worker MUST work or starve. Friedman's central claim is that the market economy we call Capitalism is NOT coercive. His claim works to the extent that we agree with his rather limited definition of coercion: no coercion exists so long as persons may choose to refuse to engage in any "particular" exchange.

The Collusion of Political and Economic Power

Friedman is justified in his concern that political power could become highly concentrated and thus interfere with human freedom. But his claim that economic power limits political power is suspect. Although one might imagine that economic and political power could be separate and thus each could limit the other's power, in reality economic power easily manifests itself as political power and vice-versa. The Fox News Channel was founded by Rupert Murdoch in 1996. His conservative political stance is well-known. In 1993 Murdoch had purchased Star TV, a Hong Kong company for $1 billion. He has clearly used his economic power to promote his political influence.

Regardless of whether one agrees with Murdoch's politics, it should be abundantly clear that political and economic powers are not distinct and that often each is been used to gain further power in the other. Jeff

Bezos (and his immediate family) is worth more than $130 billion; Bill Gates is worth about $100 billion. Who do you think has more political power, YOU or Bill Gates? Having very wealthy people does not limit the powers of political leaders; rather the wealthy trade on their economic power to buy political power. Wealthy conservatives often use their power to limit minimum wage laws, to limit worker's rights, to undercut social safety nets. This is not to say that wealthy liberals do not also use their economic power to push for their own political agendas. But the key point here is that political and economic powers are largely interchangeable and thus often *reinforce* rather than limit each other.

The Nature of Communism

When Karl Marx studied European Capitalism in the mid- 19th century, he observed many problems. He found so many problems that he thought Capitalism was ultimately doomed to failure. Capitalism would, at some point in history, be replaced with Communism. In a very real way, Communism is a criticism of the failings of Capitalism, but for our purposes we need to look at Communism itself and how it is supposed to be a fairer system. Remember that we've defined Communism as an economic system within which people get what they *need*. Thus, Communism is a system not based in rewarding people for what they do, but is, rather, a system of ensuring that as many human needs as possible are met.

The single most essential feature within Communism is the absence of private ownership of the **means of production**. People are not allowed to accumulate vast fortunes. In fact, even modest fortunes of a few million dollars would not be allowed. This limit is to be accomplished by prohibiting the ownership of what Marx calls "private property." Under Communism all factories, farms, mines, forests, major tools, and intellectual properties are owned collectively rather than privately. The abolition of "private property" under Communism often sounds frightening to people who "own" things.

Private Property is NOT Personal Property

When Marx condemns the ownership of "private" property, we need to pay close attention to how he defines that term. Regardless of how you or I might use the term in ordinary conversation, for Marx the term "private" property has a very precise definition. **Private property** includes all things that can be used to produce "profits." Wages are NOT profits. A carpenter needs tools to perform her/his job and to gain wages used to support a family. The hammers, saws, levels, drills, etc. used by a carpenter are not private property, but **personal property**. They are not private property because these tools cannot be used to produce profits; however, they can be used to help produce wages. This distinction is important. If someone were to own a factory that produces thousands of carpentry tools every year, which are to be sold for profit, then that factory is "private property" under Capitalism and would have to be collectively owned under Communism. The key distinguisher here is one of "profit" versus "wages." And, although sometimes the distinction can be hard to

see, usually the difference is obvious. My 300 square foot garden is "personal" property; the 10,000 acre farm is "private" property. My shoes are "personal" property; the factory that makes shoes is "private" property. My car is "personal" property; the factory that makes cars is "private" property. A small pile of oak boards that I may use to make a table is "personal" property; the 1,000 acre forest from which those boards come is "private" property. Under Communism, individual people and families are the true owners of their "personal" property, while "private" property would cease to exist and be owned collectively. Thus, all banks, forests, factories, farms, patents, and other items from which "profits" can be made would be owned collectively.

Who might actually be responsible for the daily management of these private properties is a question Communists find it difficult to answer. In Capitalism the owners hire the managers; in some cases, the owners assert their authority and choose to manage their properties directly. However, in Communism (where ownership is collective) a mechanism must be developed to decide on how large, productive organizations (like banks, factories, farms) are to be managed and by whom. With the idea that the people who work in factories, farms, and banks actually know a lot about them and how to maximize production, we might find a workers' council holding managerial responsibilities. However, that need not be the case. It is possible that an individual city might take managerial responsibility for factories and banks within its borders, as happens today with local fire departments. The key issue here is that any revenue generated by these local organizations would be the property of the local community, with the possibility of a wider sharing of revenue with larger political units like counties, states, and nations. What cannot happen under Communism is actually common under Capitalism: persons living hundreds or thousands of miles away owning the productive resources of an organization. Under Communism, in effect, the "stock holders" are the local community, or the workers in the organization, or the people of the nation.

Can Communism Allow for Some People to Make More Money Than Others?

Yes. Marx never says that some workers (especially more talented or harder working ones) cannot be paid more for their work. However, remember that the first order of who gets what in Communism is need, not production. So, the human needs of a community would have to be meet first, before any particularly gifted worker might be paid extra. How those "needs" might be determined is a particularly challenging problem for Communism. Under Capitalism, people are allowed to buy whatever they want with the money/resources they have. The poorest people (those lacking the sufficient resources to purchase what they need) are usually allowed to claim some additional resources (like food and shelter) beyond the limits of what they can afford. Thus, in Capitalism a "market" for "needs" exists, and who gets what is simply a matter of what any person can afford; each person (or household) makes their own decisions about what to buy. In Communism such a "market of needs" does not exist and a different mechanism for the distribution of resources has to be instituted

and managed. This can be a particularly difficult task. How does any community know who needs what? Can we trust people to simply go into a large communal warehouse and take what they need?

Beyond the basics of food, shelter, and clothing, what is a "need"? Who would decide on such things? Once the true needs of a community have been met, how might other goods be fairly distributed? Who should have access to a piano? To a super-computer? To the complete works of Plato? Communist societies tend to work best on a small scale where people know and are known by each other. In a prosperous Communist society, I certainly would not have a piano in my house (because I am totally lacking in musical talent), but I might have many books of philosophy. Would people have "spare" time in a prosperous Communist society? Would people take vacations? Not only is the distribution of resources to meet needs a challenge in Communism, but the distribution of non-need items is a challenge as well. The entire market system that plays the key role in distribution under Capitalism is absent in Communism. How this market might be replaced is, perhaps, the most difficult challenge to Communism.

The Nature of Work

In contemporary America and Europe, it is common for people to complain about their work. People often fantasize about living without having to work. Marx has an explanation for this, and surprisingly he claims that work is one of the most important and individually rewarding parts of a fully human life. However, under Capitalism work is most often **alienating:** the worker cannot identify her/himself with the work s/he performs. "Its alien character emerges clearly in the fact that as soon as no physical or other compulsion exists, labor his shunned like the plague."[7] Must all labor be alienating? According to Marx, it is an essential aspect of Capitalism that labor becomes unnatural, but work itself need not be alienating. Rather, human labor is, in its original form, the direct satisfaction of human needs. People need to eat, thus ancient humans foraged for food; people need shelter from harsh weather, thus people build houses. There is nothing un-natural about human labor. For Marx, "the productive life is the life of the species. It is life-engendering."[8] People work in order to satisfy their needs; after those needs are satisfied people will continue to labor in order to satisfy their desires—perhaps a comfortable chair, a more efficient tool, a soft bed, or some other nicety. So, for Marx, labor is not naturally oppressive; rather, labor allows humans to satisfy both their needs and (to some extent) their desires. As long as we are working to satisfy needs or to satisfy self-chosen ends, work need not oppress us. In fact, it is through work that humans discover their manifold abilities: music, science, craft, and art. In a Communist society, the population works to ensure that everyone has their needs met, and, then later, perhaps,

7. Marx, Karl & Friedrich Engels. *The Marx-Engels Reader.* W.W. Norton & Co., 2nd ed., 1978, p. 74.
8. Ibid., p. 76.

some of their desires met as well, and it is through this process of work that human beings discover their talents, refine their skills, and take pleasure in their abilities to create.

In Capitalism, labor is necessarily "alienated," and thereby oppressive. Much of the production within Capitalism is needless luxuries for the wealthy, the kinds of things that workers will never be able to enjoy: private yachts, mansions, race cars. In Communism labor provides for human needs and defines our human nature as makers. Labor may not be "fun" in a Communist society but it is supposed to be fulfilling. Under Capitalism the owners cannot pay their workers the full worth of the workers' labor; if they did there would be no profit generated for the owners, and thus no incentive to hire workers. In a Capitalist society, most workers have little or no control over what is produced or how. If you ever worked "fast food" you should know that your efforts at creativity in food production were not appreciated. In Capitalism, workers are alienated in numerous ways that tend to make work uninspiring and unpleasant. But we should remember also that Marx claims that the owners are also alienated. When large corporations fire hundreds of workers, many of the managers feel badly. They have to fire hundreds of workers who may then never find another job. Few managers actually enjoy firing workers, but they have no choice; it is merely a "business" decision. So, Marx insists that Capitalism "alienates" all members of the society, not just the workers.

The success of markets

Capitalism has the advantage of using a "**market system**" in order to determine prices and the distribution of goods, but it has the disadvantage of alienating all members of the society and of creating and maintaining a dis-advantaged class of working poor. Communism has the advantage of having no alienated class of working poor, but (lacking a market system) it has the problem of finding some process to determine prices and the distribution of resources. It should be obvious that both systems have strengths and weaknesses. The market system allows each seller to determine prices without the need for any government bureaucracy. There is a great efficiency in using a market system. If I want to buy a loaf of bread, I may choose where to buy it; if I like the taste of one brand over the other, I'm free to exercise my preference. If some seller of bread or some brand of bread does not please me, I'm free to do business elsewhere. No government bureaucracy is needed to determine prices and availability. The market efficiently decides what bread to bake, the price, the quality, and the quantity. Without a market some sort of government bureaucracy needs to be maintained in order to make these decisions, but the sheer number of production decisions that have to be made is staggering. Think of each product you can currently find at any Meijer's store. Think of all the sizes and colors available! How could any government decide in advance how to best meet the needs and wants of all consumers? It has been argued that I don't need to have 14 different sock colors to choose from. But the makers of socks have made that decision AND if they choose badly the makers of socks bear the burden of their poor choices. How could we know that fuchsia-colored socks would be wildly popular this summer?!

The problems of markets

One can easily find data on the distribution of wealth and income in America. Although the various studies may produce slightly differing numbers, a single overwhelming fact comes through: our country distributes wealth and income in a very uneven manner. In studies of wealth (as distinct from income) we find that the least prosperous 40% of American households have no wealth. The French economist Thomas Piketty sees the essential paradox of American capitalism is that fully half of the adult citizens residing in America own no capital at all.[9]

As for income, there is a widespread agreement on the basic numbers. The **median household income** in the United States (in 2018) was about $60,000. Half of American households had incomes of more than $60,000 and half had less. What is more troubling is that the poorest 25% of American households had incomes of $30,000 or less! Since there are about 128 million households, that means that nearly 30 million American households are surviving on very meager incomes. The top 1% of American households bring in $435,000 per year or more. When it comes to wealth the results are even more lop-sided. The wealthiest 0.1% of American households have the same about of wealth as the "poorest" 90%. Even mainstream business magazines admit to the problem. Here is a statement published in *Business Insider*, Oct. 23, 2017, "It's no secret that the US has an inequality problem." Here is a similar statement from *Fortune Magazine*, "There is no dispute that income inequality has been on the rise in the United States for the past four decades. The share of total income earned by the top 1 percent of families was less than 10 percent in the late 1970s but now exceeds 20 percent as of the end of 2012."[10]

Why Markets Sometimes fail

For all the advantages of markets, we need to remember that markets do not solve all economic problems and that for some social resources, markets fail to ensure anything like an adequate distribution. Housing is a notorious problem in many American cities. The average rent on the average apartment (766' sq.) in Chicago is $1,900/month. In Los Angeles the average rent is $2,380/month. The average cost of stay in the hospital is nearly $16,000–that does not include the costs of doctors services. Americans pay more for prescription drugs than any other country in the world (about half a trillion dollars—$500,000,000 this year), while drug companies are some of the most profitable companies on the planet.

Various events can undermine markets. The state of Florida (dominated by the Republican Party for the past

9. Piketty, Thomas. *Capital In The 21st Century*. Harvard University Press, 2014.
10. Chris Matthews, *Fortune Magazine*, 2014.

25 years) recognizes that sometimes markets fail. Numerous hurricanes have devastated parts of Florida again and again. People fleeing these dangerous storms often have to buy gasoline for their cars and rent hotel rooms in areas away from the storms. The storms provide opportunities for dishonest business people to raise their prices. If people need gasoline to flee from a hurricane, why not charge them $12 per gallon? If a family flees a storm and needs a hotel room for a few nights, why not charge them $500 per night (instead of the normal price of $100 per night)? With every hurricane the office of the state's Attorney General receives complaints about **price gouging**. Florida statute 501.160 makes it illegal, once a state of emergency has been declared by the governor, for businesses to raise their prices for essential items (like food, gasoline, shelter) above the average price for the previous 30 days. Why would the government of the state of Florida do such a thing? Even Milton Friedman should be able to understand that when people are fleeing from a hurricane, they are no longer "effectively free to enter or not to enter into any particular exchange." The people fleeing a hurricane MUST buy gasoline; they must buy water; they must rent a hotel room. The fundamental properties of a free exchange cease to exist when people are no longer "free" to make a particular purchase.

When Milton Friedman writes about the "market" system of capitalism, he defines the foundation of that system as a set of willing participants, one freely willing to sell and another freely deciding to buy. Numerous problems can arise in this interaction of willing buyers and willing sellers. The American Health Care system is a powerful example of a failed market. There are numerous standard textbooks that list various reasons for market failure, one form of market failure is called "**concentrated power.**" Markets for specific goods or service fail when one company gains so much power within a market that it effectively shuts out competition. Remember that for Friedman a buyer is free so long as they have the choice of selecting between various sellers. For most Americans using our health care system, there is no choice. How does victim of a terrible car crash negotiate prices with doctors on the way to the hospital? How does a person negotiate the price of a prescription drug when there is only one such drug available and no one will tell you the price? How does one even think about negotiating the price of cancer surgery? Clearly there is little or no competition in healthcare, so, according to Friedman, we are not "free" to make choices in this part of our lives; we are "coerced" in Friedman's term.

The evidence of the failure of the American healthcare system is abundant. Most studies show that Americans as a nation pay about 17% of the entire **GDP** (gross domestic product) for health care. That is, 17 cents of every dollar spent for everything Americans buy every year is spent on healthcare. When we compare that 17% figure with what people in other advanced nations pay for their healthcare we see a huge discrepancy. The advanced nations of the world pay in the range of 8–12% of their GDP. But our health system is no better. In fact, the average European lives 2 or 3 years longer than does the average American, and they pay 30-40% less for health care.

The average American lives to the age of 79.3 years. That ranks us number 31 in the world, barely ahead of Cuba and just behind Costa Rica, Chile and Slovenia. The citizens of Japan and Switzerland live the

longest—more than 83 years. The cost of health care in these countries is much less than in America. The additional billions of dollars Americans spend every year doesn't seem to help.

David Brodwin's article written for U.S. News and published in June of 2017 says,

> Health care is what economists call a "market failure." In other words, the normal logic of competition is not working. Part of the failure involves what is called "price discovery." Simply put, markets don't work well unless the buyer can discover the price of something before signing the order. How can you choose between a Tesla and a Corolla without knowing the price? Yet, it is all but impossible to find the real price for your medical services until after the service is performed, and after the insurance company (if you have coverage) has pored over the bills from the doctors, labs, hospitals and other providers.
>
> Another part of market failure involves a tremendous lack of real competition. In many parts of the country, all the hospitals are owned by one or two major chains. They charge what they want. Likewise, in many places only one or two insurance companies offer coverage for individuals and small businesses. And on top of that, drug companies have wrangled language through a prior Congress that actually prohibits Medicare and Medicaid from negotiating prices.[11]

Using Milton Friedman's definition of a free market, it should be obvious that there is no "market" for health care in the United States. Since markets can and do fail, it should be obvious that markets do not solve all our economic problems. On the other hand, markets do (in many instances) allow individual business owners and consumers to decide for themselves what prices to pay, which goods to buy, and when to walk away from bad deals.

There are various reasons that the American health care system has failed, and all of those reasons grow from the concentration of power. Medicare is forbidden by federal law from negotiating drug prices. It is forbidden by federal law to import pharmaceuticals into the US. By reducing competition, prices and profits rise. In 2016 four of the top five most profitable corporations in American were drug companies. The average large American company produced a profit of just over 10% that year; AMGEN's profits were more than 40%.[12]

Why Socialism?

Socialism as most often found today in Europe is form of **"mixed" economy** using both a market (as found

---

11. "Health Care Is a Market Failure" USNews.com, June 23, 2017. Available at: https://www.usnews.com/opinion/economic-intelligence/articles/2017-06-23/senate-obamacare-repeal- plan-ignores-market-failure-of-us-health-care

12. See AARP Bulletin, May, 2017.

in Capitalism) and (to some extent) government ownership to ensure both fair competition and that essential services be provided to all people. We need to understand that in the United States in the early 21st century we already have something like a mixed economy. Public schools, public health clinics and local fire departments all exist to ensure that all citizens are educated and healthy, and safe.

Most towns and cities have a fire department so that ALL residents (rich and poor) have protection from fires. In the early 19th century many American cities had a market of several independent fire departments. If a home owner (or business owner) wanted protection, he paid a price and placed a plaque on the exterior walls of his building naming the fire company to which he had paid a fee. If a building caught fire, all the local fire departments rushed to the fire. If there was no insurance plaque on the outside of the building, it was left to burn. That is, until, an insured building actually caught fire. However, those fire departments who had not been paid a fee, simply returned to their fire stations and let the city burn. Clearly, this is another example of the failure of a market system. And this explains why American cities no longer leave fire protection to a market. Most American cities have a community-owned fire department that protects the entire community. And most fire departments have reciprocal agreements with nearby cities to help each other in the event of major fires. We are all better off because we no longer have a market in fire protection. Socialism (the collective ownership of some capital) can be a solution to failed markets.

Bernie Sanders and "democratic socialism"

Looking closely at the economic reforms suggested by Bernie Sanders, one should be struck by the fact that it really does NOT look much like socialism. It looks like there is still a market system for determining prices; the federal government is NOT going to plan the economy—even partially.

On some issues, Bernie actually agrees with Milton Friedman. Both argue that combined economic and political power is dangerous. Both believe that markets are an essential part of any well-run economy. But Bernie sees an important aspect of the modern world that Milton misses: people with great economic power also possess great political power. Bernie recognizes that economic power skews the political system and allows a relatively small group of wealthy people to wield more political power than do the less wealthy.

If we believe that each citizen of this country is to have an equal right before the law and in the market, then allowing for great wealth is, itself, an assault on that belief. Much of Bernie's agenda is focused on ensuring greater political equality by ensuring greater economic equality. Bernie's agenda is to limit the wealth of the very wealthiest (and therefore most politically powerful) people in this country while at the same time improving (if only modestly) the economic well-being of least well-off and thereby increasing their political power. Summing up Bernie's plan for America is difficult because there is much to it, however I will try to sum it up in six major statements.

1. *Federal Income Tax rates are too low for wealthy individuals and wealthy corporations*. The highest current individual federal income tax rate in America is 37%. The maximum rate for corporate federal taxes was, until 2018, 35%; it was cut that year to 21%. However, because of various loopholes in the federal tax code, many very profitable actually have paid NO federal income tax. In the most recent year that figures are available (2018) Amazon reported profits of over $11 Billion and paid zero dollars in federal income tax. Netflix reported profits of $845 million and paid no federal income tax. There are, at least, 60 large and profitable corporations that paid no federal income tax in 2018. These include GM, Halliburton, IBM, and Eli Lilly. "Jeff Bezos, who, at this writing, is worth over $150 billion, the wealthiest person on earth. You tell me: Why should the taxpayers of this country spend billions a year subsidizing Mr. Bezos when many of his employees receive wages so low they they are forced to go on food stamps, Medicaid, or other federal programs" (Sanders, 2018, p. 246).

2. *Require Medicare to negotiate drug prices and allow the importation of drugs from other countries*. Currently, federal law prohibits Medicare from negotiating the prices they pay for pharmaceuticals. The reason for this law is that in theory Medicare is so large that it could drive prices down so low as to harm the pharmaceutical industry. But the other side of the argument holds just as well. By outlawing any negotiation of prices, the industry drives prices so high it harms Medicare's ability to pay. Federal law also prohibits Americans from importing drugs into the United States unless you are the manufacturer of those drugs. Most other countries negotiate drug prices with the manufacturers and thereby pay substantially less for prescription drugs. Bernie wants to allow Americans the same benefits enjoyed by Europeans and Canadians.

3. *Social Security has to be strengthened*. According to the Social Security Administration the poorest 33% of households run by retired Americans rely on their Social Security checks for 90% or more of their income. This number has been debated and some studies seem to show that actually only the poorest 15% of retired Americans rely on Social Security for 90% or more of their total income. Teasing out the difference between individual and household data makes things complicated. But everyone agrees that too many Americans rely too much on Social Security. Currently, Americans pay about 6.2% tax on annual salaries up to $133,000. So, for those fortunate enough to make $266,000 this year, they will only pay Social Security tax on half of their salary; this reduces such a person's effective Social Security tax rate to 3.1%. The higher your salary, the lower your effective tax rate. Bernie has proposed that we leave this salary cap in place but add a Social Security surcharge on annual salaries over $250,000.

4. *Break up the largest corporations to reinvigorate competition*. One of the most fundamental advantages of Capitalism, is that markets can force sellers to compete with one another. Competition tends to drive prices down. As we have seen, one of the most common risks to markets is the dominance of one or two powerful corporations. When competition dies, markets are skewed in favor of a small number of corporations powerful enough to set prices as they please.

5. *Increase the national minimum wage to $15/hour*. This plan would increase the economic power of the poorest working Americans. With more economic power, these Americans would have a little more

political power. Also, this would increase the consumption side of the "production/consumption" equation. No company willingly produces goods that can't be sold. So a profitable business needs customers who themselves have money to spend. While wealthy people tend to save and invest much of their income, the poor always have needs. Putting more money in to the hands of the poorest 30% of American families, would tend to push more money into consumption, which would spur business. About 42% of all American workers are currently making less than $15 per hour (which if applied to a full-time job works out to about $30,000 per year). The jobs that pay less than $15 per hour include retail sales, cashiers, food prep, office clerks, laborers, janitors and cleaners. "When you earn $9, $10, or $11 an hour, you just can't afford housing, health care, child care, transportation, decent food, other other basic necessitates of life" (Sanders, p. 246).

6. *Cap credit card interest rates at 15%.* This may seem a small thing, but many American families have credit card debt that is crushing the life out of them and reducing their ability to pay of education, health care, and save for retirement. The average credit card rate in America today is about 18%, however interest rates in general are currently (2019) at near record lows. Big banks can borrow money from the federal reserve for about 2.5%. Clearly big banks would still be able to make major profits borrowing money at 2.5% and loaning it back out at 15%.

Bernie doesn't look like a Socialist

Looking at Bernie's plan for America, it hardly even seems like real Socialism. He's not talking about workers taking over the businesses they work in. He's not talking about any sort of centralized government planning of the economy. He does talk about taking care of people's needs. Remember that even Milton Friedman says that some level of government intervention in our lives is necessary for any civilized people.

Bernie's idea of "socialism" looks more like a re-direction of lives toward our social obligations and away from our selfish wants. Too many Americans have lost both their economic and political powers. He wants to change some of the rules that perpetuate that inequality. His essential insight is that when people lack economic power they also lack political power.

Countries like Sweden, Norway, and Denmark are often called "socialist," but in reality they are not. These countries retain a set of capitalist markets, however they also provide a substantial social safety net and universal health care. These nations promote public welfare through relatively high taxes and government spending. The Prime Minister of Denmark explained that in his country,

> We have universal health coverage – you don't pay to see your doctor or go to the hospital. We have a high degree of social security. You are entitled to benefits if you lose your job, if you get sick, if you are disabled. We have one year of maternity leave, we have subsidized early childhood education and care and we ensure care for our elderly if they cannot manage on their own. We also have a strong and free

educational system. Students in institutions for higher education and university do not pay for their education, on the contrary they receive educational grants for studying.[13]

While it is true that the Scandinavian countries provide things like a generous social safety net and universal health care, an extensive welfare state is not the same thing as socialism. What Sanders and his supporters often call "**democratic socialism**" is what Europeans usually call "**social democracy**," a system in which the government aims to promote the public welfare through heavy taxation and spending, within the framework of a capitalist economy.

Warren Buffet is one of the wealthiest and most famous American investors. Recent estimates of his wealth range from $70 billion to $80 billion. In 2013 he illustrated the fundamental inequality of the American tax system. He said quite simply that the federal income tax system was unfair because his secretary paid a higher federal income tax rate than he did. For many years the income tax rate on investment income has been lower than the tax rate on salaries. Currently, a single wage earner making anywhere between $38,701 and $82,500 pays a marginal tax rate of 22%. Mr. Buffet makes his money with investments (mostly long-term capital gains) which are never taxed higher than 20%. The middle income wage earner pays a higher percentage of their income in federal income tax than does the multi-millionaire investor. Somehow the American federal income tax system taxes salaries more heavily than investment income. Middle class Americans make most of their annual income as salary. The very wealthy make most of their annual income as investment. This is simply one of a large number of tax laws put into effect to benefit the already wealthy. In a true democracy we would not enact laws that privilege the wealthy. And that seems to be what Bernie means by "socialism." He sees that gross economic inequality cannot exist without political inequality. Power manifests itself in many ways, and the essence of democracy is the dispersal of power—be that power political or economic. Concentrated wealth is a form of concentrated power and, as such, is the negation of democracy.

Thomas Jefferson wrote to a friend in 1816, "Those seeking profits, were they given total freedom, would not be the ones to trust to keep government pure and our rights secure. Indeed, it has always been those seeking wealth who were the source of corruption in government."

For deeper thought

1. How have we defined Capitalism, Communism and Socialism in this article.
2. Explain the connection between Locke's "state of nature" and Friedman's "household without need."
3. How is it that "political" and "economic" powers can provide a check on each other, according to

---

13. "Nordic Solutions and Challenges: A Danish Perspective," Kennedy School of Politics. Speech given by Lars Løkke Rasmussen, Oct. 29, 2015.

Milton Friedman?

4. Why does work suck? Can work be rewarding? How so?

5. What is the difference between European style "social democracy" and socialism?

6. In what ways are markets useful? Can markets fail? Why or why not?

7. Do economically disadvantaged people actually have less political power than do the wealthy?

8. How does Karl Marx distinguish "personal" property from "private" property? Is this a useful distinction?

CHAPTER 12: BUDDHISM

DHAMMA: WHAT THE BUDDHA TAUGHT

Douglas Sjoquist

Introduction

Many people think of Buddhism as a religion rather than a philosophy and so one might wonder why there is a chapter on Buddhism in an introduction to philosophy textbook. However, metaphysical and epistemological ideas have always been a feature of religious thinking in India (and Asia more generally). Religion and philosophy were never thought of as separate and distinct disciplines. The Buddha's teachings (called "Dhamma") addressed fundamental questions about the self, the human condition, and the nature of existence—all of which are recognized philosophical questions within the Western tradition. At the same time, the Buddha emphasized an adherence to moral practice. It might be said that intellectual philosophy and religious practice are intrinsic to the Buddha's teachings.

This chapter begins by placing the Buddha in the context of Indian history and Asian history. What were the prevailing ideas and practices in India that gave rise to Buddhism? Which philosophical/religious systems of his day did the Buddha embrace or reject? The chapter will also discuss some of the many misconceptions and misrepresentations of the Buddha's teachings. By examining these misconceptions and misrepresentations, students will gain insights into what the Buddha actually taught in contrast to what people think he taught. Tackling these mistaken views on karma, for example, serves as a useful introduction to Dhamma.

The story of Malunkyaputta and the parable of the poisoned arrow can function as a preface to the Dhamma and will provide students with an orientation to both what is important to and irrelevant in the Buddha's teachings. Inherent in the Four Noble Truths, the Noble Eightfold Path, and the Three Marks of Existence is a rational investigation of self and existence – both features of philosophical inquiry throughout human history.

The Buddha never expected his followers to blindly accept the truth of his teachings. Instead, he encouraged his followers to verify the truth of Dhamma by making their own honest observations about self and existence. This chapter lends itself to a comparative analysis between what the Buddha taught and the wisdom and methodologies of philosophers presented elsewhere in this textbook. How do the three pillars of Dhamma (wisdom, mental discipline and ethical conduct), for instance, compare with other philosophical traditions such as Roman Stoicism, American Pragmatism, or the teachings of Pythagoras? How would the Buddha view our contemporary relationship with the environment? It is hoped this chapter will stimulate these kinds of questions and contribute to a greater sensitivity to and appreciation for Buddhist thought and practice. Some of these issues will be discussed in the "philosophical afterward" at the end of the chapter.

Pali: The Original Language of Buddhism

The original written language of Buddhism is Pali. This ancient language has its origins in India and is for all practical purposes a simplified version of Sanskrit, an ancient language also from India and the original written language of Hinduism.It might be said that the relationship between Pali and Sanskrit is analogous to the relationship between Italian and Latin. That the Buddha spoke Pali is more difficult to ascertain but scholars think he probably spoke something very similar to it.Whatever the case may be, the Buddha never heard the word "Buddhism" nor did he ever see the word in writing.

It should also be noted that the teachings of the Buddha were not written down until the first century BCE. That is to say, Buddhism existed in oral form for at least four hundred years. During that time monks trained in memorization techniques memorized the Buddha's teachings. Ananda, a disciple of the Buddha, is said to have recited from memory all of the Buddha's sermons at the first Buddhist council in Rajagaha, India and Upali, another disciple of the Buddha, recited for the council all of the Buddha's 227 rules. This council was held shortly after the Buddha's death in 483 BCE. Memorization of large quantities of doctrinal information was common in India at the time of the Buddha and this was how the Buddha's teachings were passed on from generation to generation. Tradition says that at the fourth Buddhist council, convened by King Vattagamani (88-77 BCE) in Sri Lanka, the Buddha's teachings and rules were written down for the first time. In addition to the sermons and rules of the Buddha, commentaries on the Buddha's sermons by various disciples were gradually added over the years and these, too, came to be written down. In Buddhism, these written teachings, rules, and commentaries are known as the Tipitaka ("three baskets") and constitute what's called the Pali Canon. Historians believe that the teachings, rules, and commentaries were initially written at this council on palm leaves and then appropriately placed into three baskets. This is the origin of the word "tipitaka." As a written language Pali died out probably around the 14th-c. Still, scholars study this language to gain access to the original teachings of the Buddha.

Historical Context

The traditional dates for the Buddha's birth and death are 563-483 BCE, respectively. This makes him a contemporary of Mahavira (599-527 BCE), K'ung Fu-tzu (551-479 BCE), and Lao Tzu (570-517 BCE). Jainism, Confucianism, and Taoism, therefore, all came into existence in Asia about the same time as Buddhism. The area where the Buddha was born is actually now in modern-day Nepal. His mother gave birth to him in the Nepalese village of Lumbini. (Tragically, she died seven days later). Since his father was a king, he was a prince and brought up in a luxurious palace in the city of Kapilavatthu – also in Nepal. The Buddha's given name is Siddhattha Gotama (in Sanskrit, "Siddhartha Gautama"). He was born into the Sakya clan and many scholars simply refer to him by the name "Sakyamuni" (meaning "sage of the Sakya clan").

His family belonged to the nobility class. In India, this class was called the Khatiya (in Sanskrit, "Kshatriya") class. Also, the Buddha's family was Hindu. Hinduism was the dominant religion in India at the time but there was great diversity in philosophical thought among Hindus. World-views such as theism, materialism, agnosticism, determinism, nihilism, etc. all fell under the Hindu umbrella. The philosophical abstractness and the great variety in views made Hinduism virtually impossible for commoners to understand. Furthermore, many Hindu priests (called "brahmins") taught that salvation was only available to brahmins. That is, one had to be born as a brahmin (and as a male) to attain freedom from the miseries of the world and from death. Brahmins also promoted the idea that rituals were the most effective means for securing assistance from the various deities that dominated the Hindu pantheon and the only people competent to perform these rituals were the brahmins themselves. It might be said that the priests had a spiritual stranglehold on the commoner. To complicate matters even further most commoners were uneducated and could not speak, read, or write Sanskrit – the language of the brahmins. It's well-known that the priestly class had corrupted Hinduism by the time the Buddha entered the picture in India's history. India was ripe for a new world- view. The Buddha experimented with the various practices and theories associated with the Hinduism of his day and he came to the conclusion that practices such as extreme austerities and solitude were harmful and that the dominant philosophical views were erroneous. He rejected, for example, the ideas of savior beings, divine assistance, sacrifices and rituals, a creator god, and the exclusivist views of priests as means to attain salvation. (It's important to mention, perhaps, that during the time of the Buddha, the Hindu teachings represented by Upanishads had yet to reach fruition. These Upanishad teachings ushered in a Hindu reformation after the Buddha's passing).

This chapter on Buddhism will examine the central teachings of the Buddha. It is vital that students note at the onset that a distinction should be made between what the Buddha taught and what is called Buddhism. Buddhism is a religion that represents an evolution of practices and thinking that evolved over a long period of time as it migrated from one geographical area to another and split into three traditions: the Theravada, the Mahayana, and the Vajrayana.

There's great diversity in Buddhism just as there is great diversity in the various traditions of Christianity or Islam. Ronald Eyre (1929-1992), a British theater director and narrator of the television documentary series called, The Long Search, proposed a thoughtful question: "If the Buddha of Sri Lanka or India and the Buddha of Japan were to meet would they recognize each other?" The question is apropos since what the Buddha taught is quite different from what some Mahayana schools (for example, the Pure Land sects in China or Japan) taught, for example. Buddhist scholars like to distinguish what the Buddha actually taught (that is, the Dhamma) and what came to be known as Buddhism. This chapter embraces the above scholarly position and will inspect only what the Buddha taught.

The Story of Malunkyaputta

An appropriate place to begin a discussion about Dhamma is to tell the story of an exchange between a monk named Malunkyaputta and the Buddha. The scene takes place at a monastery in Savatthi, India. Malunkyaputta had been bothered by many questions that he wanted the Buddha to answer. He felt the Buddha ignored the kinds of questions that for him were important. In essence, Malunkyaputta wanted answers to questions that would be considered central issues in metaphysics or other areas of philosophy as well as dominant topics in some religions like Hinduism, Christianity, or Islam. "Is the world eternal or not eternal? Is the world finite or infinite? Is the soul the same as the body or is the soul one thing and the body another thing? After death does a Buddha exist or not exist?" He went to the Buddha and demanded that the Buddha answer these questions once and for all. Malunkyaputta declared to the Buddha that he would abandon his training and return to a normal life if these questions were not answered. The Buddha quietly listened to Malunkyaputta and then told the following parable:

> "Suppose, Malunkyaputta, a man were wounded by an arrow thickly smeared with poison and his friends and family brought a surgeon to treat him. The man would say: 'I will not let the surgeon pull out this arrow that wounded me until I know if the man who wounded me was tall, short, or middle height, dark or brown or golden-skinned, whether the man lived in a village or town or city; . . . until I know whether the bow that wounded me was a long bow or a crossbow, whether the bowstring was made of fiber, reed, sinew, hemp or bark; . . . until I know with what kind of feathers the shaft that wounded me was fitted – whether those of a vulture, a heron, a stork, a hawk, or a peacock.'" The Buddha went on in great detail and ended by saying that all these questions would still be unknown to that man and meanwhile he would die. He continued and said, "So, too, Malunkyaputta, if anyone comes to the Buddha and says he will not follow the Buddha until these questions are answered he, too, will die" (Majjhima Nikaya 63).[1]

The orientation of the Buddha's teachings is clearly illustrated in this story. The above tale shows readers that the focus of the Buddha's teachings was on the elimination of suffering and not on theory or beliefs. In other words, we as humans should concentrate our energies on removing the "arrow of suffering" rather than wasting our time on useless doctrinal speculations. Whether the universe is finite or infinite, created or non-created, etc. matters very little with regard to the realities of one's suffering. The issue of one's liberation from suffering is more important, for example, than knowing the nature of God or knowing whether the world was created or not created. The Buddha's teachings emphasize realistic solutions to human problems.

---

1. *The Majjhima Nikaya*. Nanamoli and Bodhi, translators: Wisdom Publications, 1995.

The Four Noble Truths

The Buddha was like a doctor/scientist who observed the problem with the human condition and presented a cure or solution. More precisely, he analyzed, investigated, and then offered a course of action. After his enlightenment he gave a sermon to five monks in a place called Deer Park. This sermon ("sutta") is called "The Setting in Motion of the Wheel of Teaching." At the beginning of this first sermon was the idea that we should avoid the extremes of self- indulgence on one hand and self-denial on the other. He said, "There are two extremes, monks, which must be avoided. What are these extremes? A life given to pleasures, dedicated to pleasures and lusts – this is degrading, sensual, vulgar, unworthy, and useless. And, a life given to self-torture – this is painful, unworthy, and useless." We must, said the Buddha, follow the Middle Path "which leads to insight, which leads to wisdom, which produces calm, knowledge, enlightenment, and nibbana" (Samyutta Nikaya 56).[2] The Buddha's teaching is often called the "middle path" because he taught that one should shun all extremes and instead live a life of moderation.

He then presented in this sermon what are called the **Four Noble Truths**:

1. There is suffering ("dukkha")
2. Suffering has a cause
3. Suffering can be eliminated
4. There is a way to eliminate suffering

These Four Noble Truths might be considered the essence of the Buddha's teachings on the human condition. Summarizing the Buddha's own words he said, "I teach suffering, its causes, its cessation, and the way to end suffering."

Regarding the First Noble Truth the Buddha said this: "Birth is suffering, decay is suffering, illness is suffering, death is suffering, to be united with the unpleasant is suffering, to be separated from the pleasant is suffering, not to get what one desires is suffering" (Samyutta Nikaya 56). The First Noble Truth implies that life is defective. Suffering is inevitable. To be born into the world means to experience suffering (mental, physical). To paraphrase from his book, Being and Nothingness, the French philosopher, Jean-Paul Sartre, would say "if we are born in the world we are condemned to be free." The Buddha would say, "if we are born in the world we are condemned to suffer."

The Second Noble Truth says that suffering has a cause: "It is craving (tanha) which renews being, and is accompanied by desire and lust, desire for this and that. In other words, craving for sensual pleasures, craving

---

2. *The Samyutta Nikaya*, Nanamoli and Bodhi, translators: Wisdom Publications, 1995.

to be, craving not to be" (Samyutta Nikaya 56). The Second Noble Truth identifies the cause of suffering as craving. It should be noted here that some scholars translate the Pali word tanha as "desire" but this is misleading because it implies that the Buddha taught that all desires must be eliminated to end suffering. This is simply inaccurate. Desires can be wholesome (for example, the desire to help others, the desire to be a kind person, the desire to end suffering for one's self and for others, etc.). Desires can be neutral like the desire to ride your bike or go swimming or eat when you're hungry. Other desires are unwholesome ones and these are better translated as "cravings." This translation of tanha better captures the essence of the Second Noble Truth. These unwholesome desires would include craving sensual pleasure, wealth, notoriety, and so forth.It might also include craving life and good health in the face of death or sickness, respectively. What gives rise to this craving in the first place is ignorance. We fail to see, for example, how our craving for those things, people, and activities that brings us pleasure leads to suffering.

Consider this. An object can fulfill a craving, or a person can fulfill a craving, or an activity can fulfill a craving. The fulfillment of that craving, however, brings with it attachment. In other words, we become attached to the things, the people, and the activities that brings us pleasure. This attachment is a human characteristic. It is natural. This attachment, however, gives rise to separation anxiety. We become anxious about losing the things, the people, and the activities that bring us pleasure. This anxiety is one example of suffering. It's a complicated cycle that, in Buddhism, eventually leads to rebirth. The root cause of suffering, then, is craving, which arises out of ignorance and leads to attachment that leads to suffering.

The Third Noble Truth says, "It is the complete stopping of this craving, the elimination of passions so that craving can be laid aside, given up, harbored no longer, and gotten free from" (Samyutta Nikaya 56). If the cause of suffering lies in craving, ignorance, and attachment then the elimination of suffering involves abandoning them. We don't rid ourselves of the objects, the people, or the activities; we rid ourselves of the craving and attachment to them. To eliminate suffering is to get rid of the craving and attachment and ignorance that underlie it and the Fourth Noble Truth prescribes a way to do this.

The Fourth Noble Truth says, "There is a path that leads to the cessation of suffering: it is, indeed, the Noble Eightfold Path: right views, right intentions, right speech, right actions, right livelihood, right effort, right mindfulness, and right concentration" (Samyutta Nikaya 56).

The Noble Eightfold Path

The Noble Eightfold Path (Magga) is considered the path in Buddhism. The book called the Dhammapada (literally the "teaching path") is a collection of sayings from the Buddha. It says in the book, "Of all the paths the Eightfold is the best; of truths the Four Noble are best; of mental states, detachment is best; of human

beings the illuminated one is best" (Dhammapada, 273).[3] The Noble Eightfold Path represents the middle course one must tread in life. It's the Buddhist prescription for ending suffering. That is, it is the path we must adhere to in order to avoid the extremes of self-indulgence and self-mortification and attain the awakened state like the Buddha. The NEP consists of eight principles that are usually categorized under the headings of wisdom (panna), ethics (sila), and mental discipline (samadhi). Let's look at each.

Right View simply means knowing through personal investigation and experience what suffering is, its causes, that it can be eliminated, and the way to eliminate it. It implies a correct understanding of the law of kamma, for example, and a commitment to abandoning wrong views.

Right Intention is, perhaps, the most important feature of the NEP. The Buddha asked us to consider the intent behind thought, action, and speech. In other words, whenever we act, think, or speak we should be mindful of our intent behind those thoughts, actions, and speech. Is the intent to foster harm and ill will or is the intent meant to bring about good will and harmlessness to those around you and the environment? Before you act, think, or speak ask yourself, "What is my intent?" This is important to remember since the intent of an action creates kamma.

Right Speech requires that we refrain from lying, false accusations, idle gossip, and harsh or loud talk. We should not use speech to inflame passions or incite hatred, divisiveness, or violence. Instead, speech should be quiet, compassionate, and used to create harmony in one's surroundings. Has you mother ever said, "If you don't have anything nice to say, don't say it"? If yes, she was probably a Buddhist and you didn't know it.

Right Action constitutes five principles that all Buddhists should observe. They are 1) to avoid killing, and/or harming, 2) to avoid taking that which is not given to you (that is, no stealing), 3) to avoid false speech (see above), 4) to avoid sexual misconduct (such as adultery), and finally 5) to avoid intoxication by using drugs or alcohol. It is, of course, expected of Buddhist laypeople that these precepts be observed. For Buddhist monks, these precepts and many others are more strictly prescribed. For instance, monks must remain celibate. "Having thus gone forth and possessing the monk's training and way of life, he abstains from killing living beings; with rod and weapon laid aside, gentle and kindly, he extends compassion to all living beings. Abandoning the taking of what's not given, taking only what is given and expecting only what is given. He observes celibacy, living apart from the practice of sexual intercourse." (Mijjhima Nikaya 272)

Not surprisingly, students typically ask many questions about these precepts associated with right action as it relates to them – especially to avoid killing, to avoid sexual misconduct, and to avoid intoxication. "What if a mosquito lands on my arm? Can I kill it?" "What if someone comes to my house who intends to steal or murder? Can I kill that person?" "Can my girlfriend and I have sex and still be Buddhist?" "Can my boyfriend

---

3. *The Dhammpada*, Easwaran, Eknath, translator: Nilgiri Press, 2007.

and I have a beer before or after dinner?" One way to answer these questions is to first ask, "What is the intent of my action? Will my action bring harm to another person, the environment, or myself?" Secondly, another way to address these kinds of issues for laypeople is to assess whether they conform to the practice of moderation. The Buddha warned against self-indulgence. Notice, for example, that the prohibition is against intoxication – not drinking. The prohibition is against sexual misconduct not sex.

Right Livelihood is an extension of right speech and right action into one's profession or livelihood. One must avoid deceptive practices, exploitation, violence, etc. Any occupation that brings harm to others, the environment, or oneself should be avoided. Clear examples of this would be occupations that harm or deceive others, such as engaging in human trafficking or working for an company that deploys deceptive sales tactics. One's profession should carry with it a sense of service to others or to the environment.

Right Effort means one must be resolved to cultivating wisdom, right views, right actions, etc. One cannot simply expect that things will get better or automatically improve. Suffering will not go away like magic. There must be a mental resolve to expel evil thoughts and nourish wholesome ones. There must be a mental resolve to cultivate compassion, wisdom, and generosity. Nurturing spiritual ideals requires mental discipline.

Right Mindfulness is an important feature of training the mind.It goes beyond simply having a global awareness of your surroundings at all times. It means paying attention to how one's actions, thoughts, and speech affect the environment and other people. It means paying attention to how a certain feeling or emotion has arisen in the mind. It means paying attention to how your current situation or circumstance (good or bad) has ties to past actions, words, and deeds. Right mindfulness means paying attention! Right mindfulness can mean "alertness" or "recollection" or "presence of mind," too. In the Buddha's teaching it refers to "egoless observing." One approaches the present without mental prejudice or preconception.

Right Concentration simply refers to meditation practice. Meditation practice trains the mind to focus and develop the ability to sustain a one-pointedness. It involves unifying the mind for one purpose: enlightenment or awakening.

These eight principles known as the Noble Eightfold Path are practiced concurrently and in accordance with the practice of moderation. As one can see the NEP does not represent a belief system. The eight principles represent a process, a practical guide to end suffering. These eight principles are linked and can be thought of in three categories as mentioned above: panna (Right Views and Right Intentions), sila (Right Speech, Right Action, and Right Livelihood), and samadhi (Right Effort, Right Mindfulness, and Right Concentration). The categories and their respective principles are intended to be practiced together. They are mutually supportive or interdependent. One cannot expect to conduct oneself ethically without mental discipline and mental discipline has no foundation without ethical conduct. And, both ethical conduct and mental discipline are connected with wisdom and wisdom is developed through mental discipline and ethical conduct.

The Three Marks of Existence

It's important to supplement the Four Noble Truths and the Noble Eightfold Path in this chapter with a short discussion on the Buddha's three marks of existence (tilakkhana). These characteristics do not appear as a separate and distinct sutta but they are mentioned so often in many suttas that scholars consider them essential elements in the Buddha's teachings. They should be viewed as complementary teachings to the Four Noble Truths and the Noble Eightfold Path.

The Dhammapada says this regarding these marks of existence: "All created things are transitory; those who realize this are freed from suffering. All created beings are involved in sorrow; those who realize this are freed from suffering.

All states are without self; all those who realize this are freed from suffering" (Dhammapada 277-279). In the Anguttara Nikaya (meaning "gradual sayings") the Buddha said, "Whether Tathagatas (Buddhas) arise in the world or not, it still remains a fact, a firm and necessary condition of existence, that all formations are impermanent . . . that all formations are subject to suffering . . . that all things are non-self."[4]

The first mark of existence is impermanence (anicca). That is, existence is characterized by constant change. There is no permanence in anything. There is no permanence in our thoughts, in our emotions, or in our bodies. There is no permanence in the objects of our world. All is in a constant state of change.

Ask yourself this: can I hold one particular thought or do my thoughts constantly change? Do my feelings stay the same or do they constantly change? Does my body stay the same or does it constantly change? Do the things that come to me in my experience of the world have any permanence? Clearly the answer is, "no!"All that is created is transitory. This is an existential condition.

The second mark of existence is that all beings (including animals) are subject to dissatisfaction or suffering (dukkha). We saw in the First Noble Truth that to be born in the world is to experience suffering. Part of the reason for this is that everything is subject to impermanence. Everything that changes brings with it unhappiness or distress.This is also an existential condition.

The third characteristic is, perhaps, one of the most complex and difficult principles in Buddhism to grasp for Westerners brought up in a Judeo-Christian- Islamic environment. It has even generated controversy. It's the concept called "no-self" (anatta). Normally, we think of ourselves as being the "owner" of certain features that would constitute a human being such as bodily processes, sensations, perceptions, consciousness, etc. In the Buddha's teachings, however, there is no "owner" or "self" attached to these features. The Buddhist monk,

---

4. *Anguttara Nikaya*, p. 236

teacher, and author, Ajahn Khemansanto, posed this thought: "Long ago Descartes justified the existence of self by saying 'I think, therefore I am' (cogito ergo sum). From the Buddhist point of view he almost had it correct but it should be, 'I think, therefore I think I am'. What is this thinker? Is this one who is aware separate from the thoughts, actions, feelings, or perceptions of the one who has them?" (Even Against the Wind, p.219).[5]

We utilize the word "self" for linguistic convenience to refer to the collection of perceptions, thoughts, feelings, and body but it is not a "thing" separate from those same perceptions, etc. Furthermore, those perceptions, feelings, etc. are always changing or impermanent. Simply, there is no self (that is, there is no permanent and separate entity acting as an owner of sensations, consciousness, thoughts, etc.). "Suffering exists, but not the sufferer. The act is done, but there is no doer. Peace exists, but not the one who is at peace. There is a path, but no one walks it" – Buddhaghosa (Visuddhimagga 513).[6]

Related to this third mark of existence is that Buddhists do not have a concept of a soul. Sometimes the Buddhist term anatta is interpreted as "no-soul." In Judeo-Christian-Islamic traditions it is assumed that humans have a soul that is separate from our bodies and when we die this soul continues to exist and goes to either Heaven or Hell. The Buddha put forward no such idea. There is no evidence that anything exists over and above our transitory bodies and minds that can be identified as a separate eternal entity or a "soul." This idea of no- self/no-soul is a core principle in the Buddhist ontology.

Common Misconceptions

One common misconception is that the Buddha preached a pessimistic world- view that devalues life. This wrong interpretation stems from the Buddha's principal position that if one is born into the world one can expect to experience suffering.What some critics of the Buddha's teachings ignore, however, is that he taught that suffering has causes and suffering can be eliminated.

Additionally, the Buddha taught a way to eliminate suffering! The Dhamma is not pessimistic; it's realistic. Look at any image of the Buddha. Does he have the face of a pessimist?

A second misconception is that the Buddha was a god. The Buddha never presented himself as a god nor did his disciples think of him as one. He was a human being – albeit a very unique one – who became enlightened or "awakened" to the realities of life and then taught people how to discover their own awakening and overcome suffering associated with these realities. What are the realities of human life? According to the Buddha it

---

5. *Even Against the Wind*, Khemasanto: Dhammasala Forest Monastery, 2000.

6. *Visuddhimagga*, Nanamoli, translator: Buddhist Publication Society, 1991.

is, first and foremost, that although we all want *lasting* happiness and pleasure, we find only frustration, disappointment, and impermanence in that which brings us pleasure and happiness. The word "Buddha" means "one who is awake" (to the nature, causes, and elimination of suffering).

A third misconception is that the Buddha taught reincarnation. There is no notion of reincarnation in the Buddha's teachings like there is in Hinduism. There is no soul that migrates from one lifetime to the next. What the Buddha taught was the notion of rebirth. Life does indeed continue after death but it is kammic/karmic tendencies rather a soul that migrates from one life to another.

It seems there is great misunderstanding of the notion of kamma ("karma" in Sanskrit), as well. The Buddhist interpretation of kamma is distinguishable from the Hindu notion of karma. In Hinduism, karma can mean "action" or the consequences of a physical or mental action. It can also mean the sum of all consequences of actions. Furthermore, Hindus think of karma as a cause-and-effect relationship operative in human behavior. Buddhism generally accepts these ideas about karma found in Hinduism. However, in Buddhism the link holding the universal law of cause-and-effect together is intention. Actions only produce results under certain circumstances. In other words, the effect of an action is not primarily determined by the act itself but rather by the intent of the action. It is the conscious intention of actions that causes kammic/karmic effects to arise. For example, if you unintentionally run over a squirrel with your car on the way to school it does not necessarily mean that you'll accumulate "bad karma" as a result. There has to be some conscious intent behind the action.

In today's "New Age" thinking people often embrace the idea that a person who suffers from a certain fate (for example, someone who has cancer, someone who was murdered, someone with a birth defect, etc.) deserved that fate due to their actions in the past – perhaps even lifetimes ago. The Buddha warned his monks not to fall into that wrong thinking. Humans are conditioned not only by their kamma but also by genetics, their environment, physical laws, and the mind, as well. Most importantly, it is the intention of an action that causes a certain effect in the future.

This brings up two other points regarding the Buddhist notion of kamma/karma. John Lennon sang a song called "Instant Karma." This idea of kamma/karma is also erroneous. The effect of an action is not instantaneous! In other words, an action and the effect of an action cannot happen simultaneously. The consequence of an action may not manifest until several days, years, or even lifetimes later. Secondly, the effect of an action is not necessarily an effect in kind. In some religions you find phrases like "an eye for an eye and a tooth for a tooth" or "as you sow, so shall you reap." This, too, is erroneous thinking from the perspective of Buddhism. For example, if you cheat on your girlfriend the consequence of that action might be that she leaves you forever instead of reciprocating by cheating on you while in the relationship. For every intentive action there is a reaction but not necessarily a reaction in kind or in proximity to the action itself.

There are many misrepresentations and misinterpretations regarding Buddhist concepts. This applies to Buddhist art, as well. Many people have seen the imagery below. This is not the Buddha. This is an image of Pu-

tai, a Chinese Buddhist monk associated with Ch'an Buddhism. He is associated with prosperity and longevity and is a well-known figure in Chinese popular culture. This imagery dates back to about the 10th-c. The image is commonly referred to as the "laughing Buddha" or the "fat Buddha." This is not, however, the historical Buddha.

(Image: Douglas Sjoquist)

There are other misconceptions about Buddhism and what the Buddha taught (for example, all Buddhists are vegetarians, most Buddhists live in India, all Buddhists practice meditation) but the above represent the most commonly held ones.

Concluding Remarks

The Buddha did not create a theology, a cosmology, a cosmogony, or eschatology. There are no divinely revealed scriptures regarding what he taught. He did not develop a liturgy or prescribe any rituals. In other words, the Dhamma is not a religion in the way that many of think about religion. Rather, it is a path – a path that suggests mental discipline, living a life of moderation, being mindful of our intent behind thoughts, words, and actions, and practicing generosity and kindness will lead to insights on the human condition and eventually an awakened state. This awakened state in Buddhism is referred to as "enlightenment" (bodhi). What exactly is one awakened to? One is awakened to the nature of suffering, the cause of suffering, the realization that suffering can be eliminated, and the complete understanding that there is a way to eliminate it. Ultimately, the goal is nibbana (literally, "to be extinguished") – the highest kind of enlightenment.It says in the "Udana" (the third subdivision of the *Khuddaka Nikaya*): "There is, monks, that state where there is no earth, no water, no fire, no air; no base consisting of the infinity of space, no base consisting of the infinity of consciousness, no base consisting of nothingness, no base consisting of neither perception nor non-perception; neither this world nor the next nor both; neither sun nor moon. Here, monks, I say there is no coming, no going, no staying, no decreasing, no uprising, no fixed, no moveable, it has no support. Just this is the end of dukkha" (Udana 8.1).[7]

Study questions

1. It seems that Buddhism arrived in Indian history at an opportune time. Describe some of the characteristics of Hinduism – especially the Brahmin class – that made India eager to embrace a new teacher and his teachings.

2. What is the name for the Buddha's teachings? Why do scholars (and some Buddhists) make a distinction between what the Buddha taught and Buddhism?Do you think such a distinction has merit?

3. Identify at least three common misconceptions about Buddhism. (NOTE: Include in your answer the concept of kamma and two other misunderstandings about Buddhism commonly found in the West). How do you suppose these misconceptions arise and why do you suppose they get perpetuated? Can the

---

7. The Khuddaka Nikaya, Thanissaro, translator: Metta Forest Monastery, 2005.

same be said about the teachings of Jesus and/or Muhammad?

4. What message was the Buddha trying to convey in telling Malunkyaputta the parable of the poisoned arrow?

5. Buddhism is often referred to as "the Middle Path." Why?

6. We all enjoy the simple pleasures in life like having coffee in the morning or putting on headphones and listening to our favorite music in the evening. Illustrate how engaging in a simple pleasure might lead to a form of suffering.

7. According to the Buddha's teachings, if various forms of suffering are tied to ignorance, craving, and attachment how, then, is this suffering to be eliminated?

8. Arrange each of the principles associated with the Buddha's Noble Eightfold Path according to the following categories: Wisdom, Mental Discipline, and Ethical Conduct. Illustrate how these principles (and categories) are mutually supportive.

9. "Suffering exists, but not the sufferer. The act is done, but there is no doer. Peace exists, but not the one who is at peace. There is a path, but no one walks it." – Buddhaghosa (Visuddhimagga 513). What Buddhist concept does Buddhagosa refer to in the above quotation? Explain.

10. What elements in Western religious traditions are not found in the Buddha's Dhamma? What elements in Western philosophical thought are not found in the Dhamma? Does the absence of these elements imply that Dhamma is neither a religion nor a philosophy?

Philosophical Afterward (Matthew Van Cleave)

Many of the chapters in this textbook address theoretical and conceptual problems. The problem of other minds, external world skepticism, the problem of free will and determinism—none of these issues concern the practical matter of how to live well. Rather, they are more intellectual than practical. Although much of the western philosophical tradition focuses on these intellectual and conceptual issues, the concern with practical matters, such as how to live well, has always concerned philosophers, as well. Going all the way back to ancient Greece, Socrates was first and foremost concerned with how to live well—with "ethics" in the broadest sense of that term. Today, ethics remains a flourishing area within the discipline of philosophy, alongside metaphysics, epistemology, and logic. Although contemporary philosophers have tended to intellectualize much of the discipline of ethics, there remains ongoing interest in older traditions that emphasize the practical and therapeutic over the intellectual.

Buddhism, and specifically the Dhamma, is clearly philosophical in this sense. As the story of Malunkyaputta suggests, we often don't need a subtle *intellectual* understanding of things in order to reap *practical* or *therapeutic* benefits. Indeed, Malunkyaputta's questions only inhibit his reaping the therapeutic benefits of the surgery. One philosophical tradition that exemplifies this therapeutic conception of philosophy is Stoicism.

Since Stoicism is interestingly similar to the Buddha's teaching about how desire leads to suffering and that living well involves taming our desires, it would be instructive to consider this similarity.

Stoicism began in ancient Greece and was imported to ancient Rome, where it was developed in the hands of influential Roman statesmen and emperors, including Cicero, Seneca, Epictetus, and Marcus Aurelius.[8] As we have seen in this chapter, the second noble truth identifies the cause of suffering with faulty desires and suggests that in order to live better we should constrain not only *what* we desire, but also the *way* we desire things. The third noble truth suggests that we can (and should) learn to desire without getting attached to things. Stoicism makes a very similar claim about desire: that we should only desire those things which are under our control. Stoics had an interesting way of reconciling free will with determinism. On the one hand, Stoics believed that everything that happened was fated (by god/nature) to happen and it could not have happened otherwise (this is determinism). On the other hand, they believed that if we could come to desire whatever it was that happened, then whatever happened is something that we would want and in this sense we would have freedom.[9] In contrast, we lack freedom when our desires are in conflict with what happens in the world. Thus, according to the Stoics, if we want to increase our freedom, we should learn to be accept with equanimity whatever happens. We should focus our energy on what we can control and although we cannot control what happens in the world, we can control *our reactions to* what happens in the world (our own minds). As Epictetus said, "Don't demand that things happen as you wish, but wish that they happen as they do happen, and you will go on well."[10] The Stoics called this state of mind of calm acceptance of whatever happens **apatheia**. Indeed, some Stoics such as Epictetus pushed this acceptance to the extreme:

> With regard to whatever objects give you delight, are useful, or are deeply loved, remember to tell yourself of what general nature they are, beginning from the most insignificant things. If, for example, you are fond of a specific ceramic cup, remind yourself that it is only ceramic cups in general of which you are fond. Then, if it breaks, you will not be disturbed. If you kiss your child, or your wife, *say that you only kiss things which are human, and thus you will not be disturbed if either of them dies*.[11]

Thus the third noble truth and the Stoic notion of apatheia seem to share the idea that in order to live well, we should tame and control our desires so that we do not become attached to things that are not within our control and that we could lose.

---

8. Juczak, Paul M. 2014. *Worldy and Unworldly Philosophies: From the First Philosophers to the 15th Century*. Open textbook available through author: jurczap@lcc.edu

9. Stoicism thus takes a "compatibilist" position regarding the problem of free will and determinism. See the chapter on free will and determinism in this textbook.

10. *Epictetus*, Enchiridion, 8. Available at http://classics.mit.edu/Epictetus/epicench.html

11. *Epictetus*, Enchiridion, 3 (emphasis mine). Available at http://classics.mit.edu/Epictetus/epicench.html

This is both a serious but contentious claim about what human beings need to do in order to live well. Not everyone agrees that in order to live well we should narrow the scope and nature of what we desire, but this same issue is one that arises in other places within philosophy's long history. Martha Nussbaum has argued that there was a longstanding debate on exactly this issue between the ancient Greek tragedians (like Euripides, Sophocles and Aeschylus), on the one hand, and Socrates and Plato, on the other. The tragedians, as Nussbaum reads them, are trying to show us that because of the scope and complexity of what humans desire, we will inevitably have to face tragic choices in life. These tragic situations cannot be avoided. Socrates' response to the tragedians, as Nussbaum sees it, was to grant that they were correct in seeing that most human lives would involve tragedies—conflicts between things that we equally deeply value—but to suggest that we could escape the tragedy by narrowing the range of what we desire. For example, if we could reduce everything we valued to one kind of value, then there wouldn't be any conflicts between incommensurable values (such as loyalty to family versus loyalty to state or religion). Consider love: if we could "ascend" to valuing only abstract beauty and not particular beautiful people, then there is no risk of loss or hurt. (This is exactly what Nussbaum argues is Socrates' position in Plato's *Symposium* and it also sounds remarkably similar to the Epictetus quote above.) Nussbaum argues that Plato is well aware of the fact that ascending to this kind of abstraction involves giving up something that is a deep part of our humanity and that safeguarding ourselves from loss in this way isn't worth the price of making ourselves less human.[12] Perhaps there is a similar philosophical debate to be had regarding the viability of the third noble truth and of the Stoic ideal of apatheia.

There is another clear philosophical connection to be made with the Buddhist concept of *anatta* (no-self). Claims about what "the self" is connect with the traditional philosophical issue of personal identity (see the chapter in this textbook). One of the key questions that philosophers have asked about personal identity is how we can persist through time. Consider the fact that who you are now and who you were when you were 7 years old are radically different in almost every way. So, in what sense are we talking about the same person in these two instances? This is what philosophers have labelled "the persistence question" (see the personal identity chapter in this textbook). Some philosophers have taken the view that there is, in fact, nothing that persists through time and thus that we do not have a persisting self. Such philosophers seem to be in alignment with the Buddhist concept of *anatta*. One Buddhist text in which the doctrine of anatta is clearly stated (and that reads very similarly to philosophical writing about personal identity) is *The Questions of King Milinda* in which the Buddhist sage, Nāgasena, explains to King Milinda that just as a chariot is nothing in additions to all its parts, so the self is nothing—no extra thing—in addition to all of our parts. I quote that text here at length.

Now Milinda the king went up to where the venerable Nâgasena was, and addressed him with the greetings and compliments of friendship and courtesy, and took his seat respectfully apart. And Nâgasena reciprocated his

---

12. Martha Nussbaum argues all of the preceding points in her magisterial work, *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy*. Cambridge University Press, 2001.

courtesy, so that the heart of the king was propitiated. And Milinda began by asking, 'How is your Reverence known, and what, Sir, is your name?'

'I am known as Nâgasena, O king, and it is by that name that my brethren in the faith address me. But although parents, O king, give such a name as Nâgasena, or Sûrasena, or Vîrasena, or Sîhasena, yet this, Sire,– Nâgasena and so on–is only a generally understood term, a designation in common use. For there is no permanent individuality (no soul) involved in the matter.'

Then Milinda called upon the Yonakas and the brethren to witness: 'This Nâgasena says there is no permanent individuality (no soul) implied in his name. Is it now even possible to approve him in that?' And turning to Nâgasena, he said: 'If, most reverend Nâgasena, there be no permanent individuality (no soul) involved in the matter, who is it, pray, who gives to you members of the Order your robes and food and lodging and necessaries for the sick? Who is it who enjoys such things when given?

> Who is it who lives a life of righteousness? Who is it who devotes himself to meditation? Who is it who attains to the goal of the Excellent Way, to the Nirvâna of Arahatship? And who is it who destroys living creatures? who is it who takes what is not his own? who is it who lives an evil life of worldly lusts, who speaks lies, who drinks strong drink, who (in a word) commits any one of the five sins which work out their bitter fruit even in this life? If that be so there is neither merit nor demerit; there is neither doer nor causer of good or evil deeds; there is neither fruit nor result of good or evil Karma. If, most reverend Nâgasena, we are to think that were a man to kill you there would be no murder, then it follows that there are no real masters or teachers in your Order, and that your ordinations are void.–You tell me that your brethren in the Order are in the habit of addressing you as Nâgasena. Now what is that Nâgasena?

Do you mean to say that the hair is Nâgasena?' 'I don't say that, great king.'

'Or the hairs on the body, perhaps?' 'Certainly not.'

'Or is it the nails, the teeth, the skin, the flesh, the nerves, the bones, the marrow, the kidneys, the heart, the liver, the abdomen, the spleen, the lungs, the larger intestines, the lower intestines, the stomach, the fæces, the bile, the phlegm, the pus, the blood, the sweat, the fat, the tears, the serum, the saliva, the mucus, the oil that lubricates the joints, the urine, or the brain, or any or all of these, that is Nâgasena?'

And to each of these he answered no.

'Is it the outward form then (Rûpa) that is Nâgasena, or the sensations (Vedanâ), or the ideas (Saññâ), or the confections (the constituent elements of character, Samkhârâ), or the consciousness (Vigññâna), that is Nâgasena?'

And to each of these also he answered no.

'Then is it all these Skandhas combined that are Nâgasena?' 'No! great king.'

'But is there anything outside the five Skandhas that is Nâgasena?' And still he answered no.

'Then thus, ask as I may, I can discover no Nâgasena. Nâgasena is a mere empty sound. Who then is the Nâgasena that we see before us? It is a falsehood that your reverence has spoken, an untruth!'

And the venerable Nâgasena said to Milinda the king: 'You, Sire, have been brought up in great luxury, as beseems your noble birth. If you were to walk this dry weather on the hot and sandy ground, trampling under foot the gritty, gravelly grains of the hard sand, your feet would hurt you. And as your body would be in pain, your mind would be disturbed, and you would experience a sense of bodily suffering. How then did you come, on foot, or in a chariot?'

'I did not come, Sir, on foot. I came in a carriage.'

'Then if you came, Sire, in a carriage, explain to me what that is. Is it the pole that is the chariot?'

'I did not say that.'

'Is it the axle that is the chariot?' 'Certainly not.'

'Is it the wheels, or the framework, or the ropes, or the yoke, or the spokes of the wheels, or the goad, that are the chariot?'

And to all these he still answered no.

'Then is it all these parts of it that are the chariot?' 'No, Sir.'

'But is there anything outside them that is the chariot?' And still he answered no.

'Then thus, ask as I may, I can discover no chariot. Chariot is a mere empty sound. What then is the chariot you say you came in? It is a falsehood that your Majesty has spoken, an untruth! There is no such thing as a chariot! You are king over all India, a mighty monarch. Of whom then are you afraid that you speak untruth? And he called upon the Yonakas and the brethren to witness, saying: 'Milinda the king here has said that he came by carriage. But when asked in that case to explain what the carriage was, he is unable to establish what he averred. Is it, forsooth, possible to approve him in that?'

When he had thus spoken the five hundred Yonakas shouted their applause, and said to the king: Now let your Majesty get out of that if you can?'

And Milinda the king replied to Nâgasena, and said: 'I have spoken no untruth, reverend Sir. It is on account of its having all these things—the pole, and the axle, the wheels, and the framework, the ropes, the yoke, the

spokes, and the goad–that it comes under the generally understood term, the designation in common use, of "chariot."'

'Very good! Your Majesty has rightly grasped the meaning of "chariot." And just even so it is on account of all those things you questioned me about–the thirty-two kinds of organic matter in a human body, and the five constituent elements of being–that I come under the generally understood term, the designation in common use, of "Nâgasena."[13]

I will end this philosophical afterward with a question I had when reading this chapter. It is clear that the Buddha put forward the four noble truths and eightfold path and that he thought that these were crucial for humans to understand in order to live well. But how did the Buddha know this? This is an epistemological question; I am asking why we should believe that what the Buddha said was true. This question is especially pressing given that the Buddha doesn't really try to give any arguments or reasoning for why what he said was true. What does seem to be clear from this chapter is that the Buddha discovered these truths from experience—from actually living it himself. It is here perhaps that we can tie the Buddha's way of answering this question to an existing American philosophical tradition: pragmatism. American pragmatists (the big three being C.S. Pierce, William James, and John Dewey) suggested that what we mean by truth is simply what we can establish through experience. In short, if x works well, then x is true. Whatever might be our ultimate judgment about pragmatic theories of truth, there is another question that seems to me more pressing. The question is: Even if we grant that living in this way (following "the path") worked for the Buddha, why should that mean it will thereby work for me? Of course, if I try it and find it works, then I have my answer. But why should I invest the time an energy into this particular path when there are so many other "paths" that other traditions think I should follow? These are questions that anyone who thinks deeply about what it means for humans to live well cannot avoid asking. Buddhism enters its answer into a mélange of other answers that have been offered by other philosophical and religious traditions. The role of philosophy is to sort through this mélange of answers for the truth.

---

13. Note that *The Questions of King Milinda* is not part of Dhamma, is not a sermon of the Buddha, is not found in Thai or Sri Lankan translations of Pali Canon, and is not in the sutras of the Pali Canon in Theravada tradition (except Burma). Full text of *The Questions of King Milinda* is available here: http://www.sacred-texts.com/bud/sbe35/sbe3504.htm

CHAPTER 13: AESTHETICS

ISSUES IN AESTHETICS: HOW TO JUDGE ART?

Christopher Schneck

We've all had the argument with our good friends about whether our favorite TV show is good. Say you watched the most recent season of *Game of Thrones* and want to argue with your friends who say it's a terrible season – how would you go about doing so? Do you argue with them on the basis that you like the CGI characters, the writing, the allegories the story expresses (if it does), what? It can feel really difficult to discuss these things because they are so close to us. It also often feels difficult because we lack the concepts to effectively convince another person why something we like is something they should also like. Sometimes we make appeals to morals, appeals to objectivity, or even force. The idea that others should like things we like has its roots in an idea of taste – if we like something, it has a quality everyone could like.

Within philosophy, **aesthetics** is the study of beauty. This study is done especially in art, though consideration of beauty solely within the domain of art was only a recent development of philosophical considerations of beauty. While aesthetics is mostly preoccupied with art, there are other areas of experience where we also care about beauty, such as in nature or mathematics. Rather than a moral claim about what others should be like, judgments of taste appeal to beauty, a quality of the object others can also appreciate.

This chapter will deal primarily with judgments of taste. You might wonder why we don't begin with the artworks themselves and discuss whether something is good art or not. This is a good question to ask – the difficulty lies in what presuppositions we have about what we like and how we can discuss those things we like. What does it mean to like *CSI: Miami* more than *Deadwood*, or does it mean anything at all? Or, what if we like going to the Opera rather than watching *Deadwood*, is one more artistic than the other? What you will see is that many of these distinctions rely first on our conception of beauty and after that on how we distinguish our taste from others'.

To judge different works of art means to rely on a few basic concepts that describe our experience, especially the concepts of **beauty** and **taste**. A thing is beautiful if it attracts us and gives us pleasure, and we exercise **taste** in identifying the things that please us and differentiating them from other things. This means that we're not just talking about cathedrals or supermodels, but a more general use of beauty in which it is always the ultimate object of our desires. The beautiful object is the object that gives us pleasure. To exercise our taste and judge art is to judge objects that give us pleasure, and therefore to make judgments about beauty. Though we now question whether artworks are primarily beautiful objects this is still done by making judgments about what we like, we may just have a different name for what it is we like. These are judgments of taste.

As a specific subdivision of philosophical inquiry, aesthetics only dates to the 18th century. The emergence of a separate discipline of philosophy dedicated to art developed out of the romantic period, and was especially important as a reaction to the enlightenment philosophy of the time. The term 'aesthetics' was chosen for the philosophy of art to emphasize that not all knowledge is scientific or factual, but that there are independent ways of knowing the world through the experience of art.

Instead, some knowledge is aesthetic and pertains to our feelings about the world. You might consider, for instance, what Homer's *Odyssey* teaches you about being an individual. Most of what you learn from classics like Homer aren't given to you in the form of deductions about the concept of an individual. Instead, through learning the motifs of the story and applying them to your own experiences, you learn how to be an individual without the need for any deductive knowledge about the concept 'individual'.
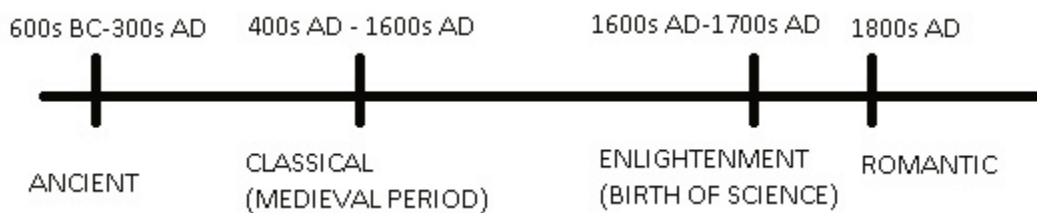
Prior to this division between romantic and enlightenment philosophy about artistic knowledge in the 18th century, the ancient Greeks considered beauty to be the ultimate value of cosmos. The **romantics** stressed non-scientific qualities of art, like feeling and creativity. The **enlightenment** thinkers stressed the proportion, order, and complexity of artwork as aspects of its technical and mathematical perfection. The enlightenment thinkers were largely carrying on the traditions of ancient and classical aesthetics. For ancient Greek thinkers like Plato, beauty was an effect of order, which meant that properly rational things should be beautiful. For example, the use of perfect proportions in music has long been heralded as a standard of beauty, such as in the case of Bach's counterpoint or Stravinsky's polyrhythmia. In these examples different notes work together to form a harmony which creates a beautiful sound, and this is attributed in many cases to the proportion and mathematical elegance of the written composition.

Leading up to the 18th century, most philosophers agreed with this Platonic model of beauty. Ideas about what was beautiful remained pretty consistent from the ancient Greeks all the way through the Enlightenment period in the graph above. The more rational something was, the more beautiful it was. In this way, beauty was an aspect of the truth of the thing under consideration. This is very different from the romantic idea of beauty. For the romantics, a beautiful thing wasn't something ordered and rational. In fact, it often seemed to be the opposite – beautiful things make you feel their beauty, they are inexact and messy, and they often depend on your own subjective experience.

The historical difference between the enlightenment philosophers' objective view of beauty and the romantic philosophers' subjective view of beauty isn't just a historical point. That difference is a good fault line to show

what the typical differences are in philosophical conceptions of beauty. The division between these groups of philosophers motivates many discussions within aesthetics distinct from simply the idea of beauty, as well. As you will see in this chapter, many of the discussions of artistic value depend on this key difference between subjective and objective value, and how it maps onto other considerations within aesthetics. Examples include:

- is beauty natural or manmade?
- is art representative, expressionistic, or reflective?
- does art have objective meaning or does it simply exist for personal pleasure?
- is the meaning of an artwork determined by the artist's intention or is it determined by its cultural and historical context?

The Idea of Beauty

Before we consider artistic beauty specifically, we should understand what makes something beautiful. Some important considerations for understanding beauty are: the relationship between form and content, whether beauty is objective or subjective, and whether beauty is a natural quality or made by humans.

While it's natural to want to consider either the artwork in itself or our individual experience of it, we should first clarify some basic aspects of the artwork. When we try to communicate why something is beautiful to us, we often have a personal feeling about the content of the work. Our favorite TV show might really speak to us: our ambitions, our identity, or it might mirror some representation of how we see the world. We don't often think about how these representations are presented to us as part of the pleasure we experience. In other words, we don't often think about the form of the art we like, just the stuff we find pleasurable in the artwork.

However, **form** plays a big role in the presentation of an artwork. When we refer to the form of an artwork, we mean the composition or techniques used to create the artwork. Form is a manner of portraying the content of an artwork. Content refers to the artwork's subject matter or its meaning. So painting is a form of artwork, as opposed to other forms of artwork like sculpture or film. We can also say that different ways of applying paint—spray paint for a mural vs. traditional brushstrokes for a Renaissance painting—are different forms of painting.

Van Gogh's paintings (here Landscape with Wheat Sheaves and a Rising Moon) illustrates the difference in texture different techniques can create. The form of painting – Van Gogh's brushstrokes – has an immediate effect on how you receive the content. The content, the wheat fields, appear to take on a different meaning when portrayed in Van Gogh's painting than in a traditional oil landscape painting.

For some thinkers, beauty is a product of the perfection of an artwork's form. A good example of this is the classical sculptures of ancient Greece, or the previously mentioned composers' music. For these forms of art, we often feel they express a perfect harmony of parts, and that what the sculpture or the music is about is relatively meaningless. In this way, we judge these artworks according to their form, rather than their content. Stravinsky's **polyrhythmia** could be his attempt to portray the complexity of life, or it could be his attempt to represent a Greek myth, or it could simply be an attempt to create a beautiful harmony. In any of these cases, though, it doesn't change the beauty of his musical composition. The artwork is formally beautiful, though its content may or may not be beautiful.

For many of us, it's hard to think about something being beautiful in this way. This formalist view of beauty

grew out of ancient Greek philosophy, and was heavily influential on considerations of beauty up through the Renaissance, but it doesn't seem to hold much sway over us nowadays. Often, it feels like we are drawn to artworks because of the content represented in them. This could mean that we like feeling represented in an artwork, that it gives us an identity or that it speaks to our personal commitments, or any number of other psychological reasons. While some artworks seems to be beautiful simply by virtue of their perfection of the form of their art, others seem to only be beautiful by virtue of how they personally move us. This distinction in art, the distinction between form and content, will also map very well onto another dichotomy within aesthetics: whether beauty is subjective or objective in nature. This problem often breaks down like this: if beauty is a product of the form of an artwork, it is an objective quality of the artwork. Thus, if beauty is a product of the content of an artwork, it is a subjective quality we experience about the artwork. Though this distinction is way too cut-and-dry, we'll see that it does capture much of the spirit of our upcoming considerations on whether beauty resides in an artwork or in the person viewing the artwork.

Objective views on beauty

The first view we will consider is the **objective view of beauty**: beauty is a quality of the object. In this view, beauty is a quality that exists independent of any individual person. This means that 'beauty' is something that naturally exists, independent of any individual person's desires. A beautiful thing expresses beauty insofar it meets objective criteria that allow us to put it in the class of 'beautiful' things. The renaissance painters we discussed earlier are a good example for this view, since their paintings are beautiful insofar as they express beautiful proportions or forms. These paintings don't move us the same way our favorite comics or movies do, but we can see that they are beautiful, even if it has little to do with our own interests.

Often, objective beauty is an aspect of the form of the object, rather than the content. People who subscribe to this view are called **formalist**. Again, in the renaissance painters the subjects of the paintings (men, women, children, goats, apples, etc.) are not themselves beautiful. The content in these paintings, their subjects, matter very little. We would hardly compare supermodels today with the Mona Lisa! Rather, the proportion of the paintings approaches a mathematical and formal perfection that those painters called 'beautiful'.

This brings up an important question about the view that beauty is objective: is beauty a product of the order or proportion of the thing? This is a difficult view to sign on with. This would mean that the content of the artwork doesn't matter: neither the characters of the novels we read, nor the superheroes in our favorite comics, nor the subjects portrayed in a painting, nor even the lyrics in our favorite songs! Even though these aspects of our favorite art seem like the most important part, the formalist would argue that they do not meaningfully contribute to art.

It's also possible that objective beauty isn't wholly dependent on the form of the art. With increasing complexity, something becomes increasingly beautiful. A good example of this is a Gothic cathedral – many

of the striking features of these churches are due to their overwhelming complexity. This view is often tied to scientific or mathematical concepts of beauty, as well. The flying buttress on the aforementioned gothic church is beautiful precisely because of the complexity of its architecture and its mathematical dimensions. Shortly, when we discuss the sublime, we will see that this is a major factor in determining whether things are beautiful because they are too complex for us to understand in everyday terms.
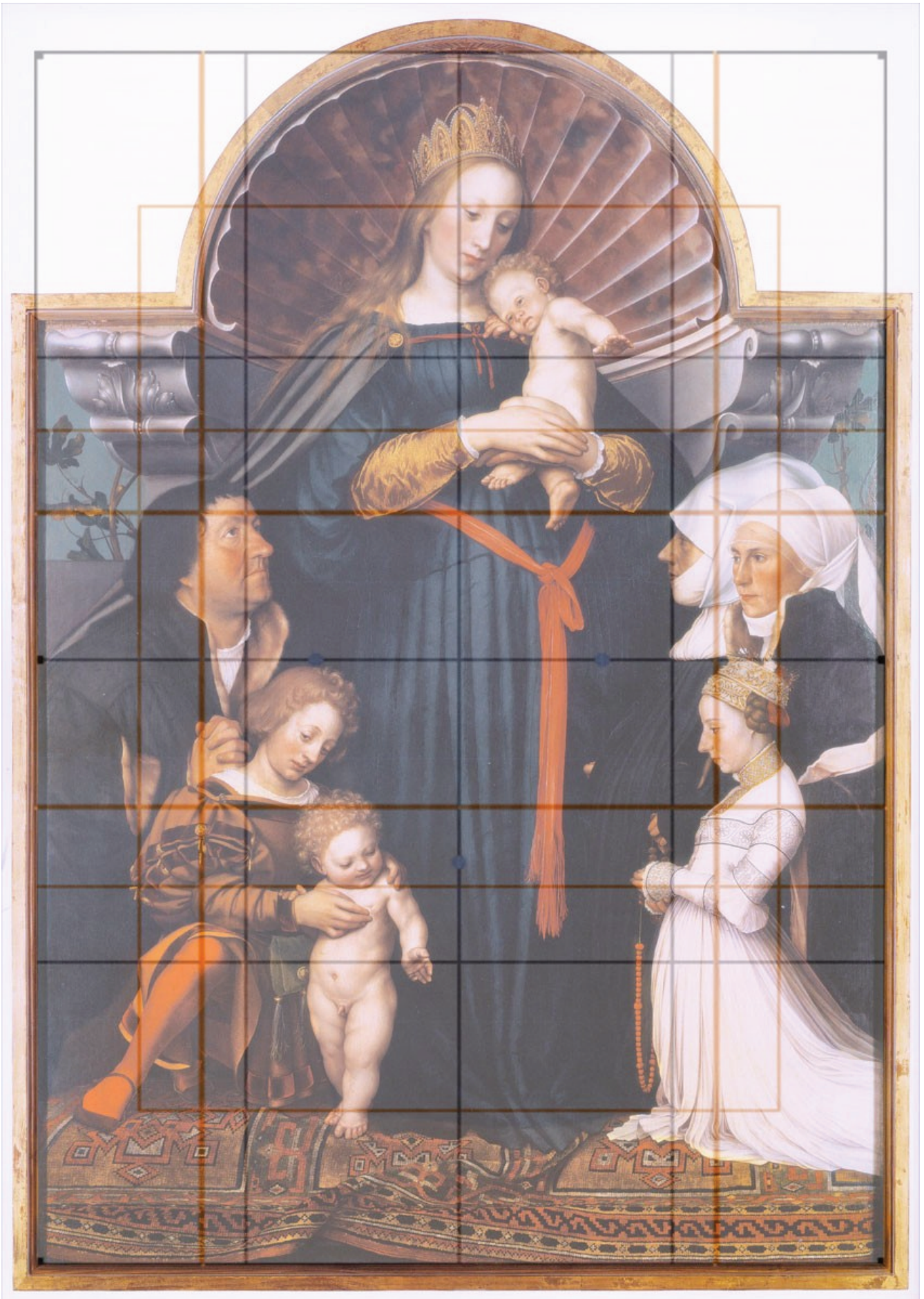
You can see in the Darmstadt Madonna that every part of the image is constructed to fit into sectors according to a pre-planned proportion, according to those three main vertical lines that trisect the image. In this painting, the Madonna is less important than the proportion of the image.

Our last consideration for objective beauty should be whether there is beauty in itself. Broadly, the question we should ask is "Is beauty a quality independent of the thing that expresses it?" If we answer "yes", then beauty must be something that exists independent of beautiful things. While it may be tempting to say that we can just generalize from specific beautiful objects to beauty in general, this would not give us an objective beauty. If beauty is an objective quality, there must be some independent standard to measure specific beautiful things by; we should have some way of relating the Mona Lisa and a Gothic cathedral and measuring whether one is more beautiful than the other.

Subjective views on beauty

As we just discussed, much of history has consigned beauty to an objective existence separate from our individual experience. This may seem, however, totally alien to your own personal experience. Lots of us think of art as something that has a personal effect on us, or pertains to our individual taste. This is because we've become used to viewing art as a subjective experience. If beauty is **subjective**, it must mean that beauty is a quality of the individual's experience rather than a quality of the artwork. There are a few different ways we can talk about subjective beauty: as an aspect of taste, pleasure, and entertainment. Since these concepts rely on the individual's subjectivity, they are very different from our previous considerations of the object itself via its form, order, and complexity.

As we previously discussed, the objective view of beauty was largely dominant from the ancient through late medieval period. With the enlightenment of the 17th century came an increasing emphasis on the individual person and their experience, rather than how that individual fit into the objective picture of the world. Accordingly, judgments of taste became one of the main ways to talk about pleasure, and therefore beauty.

While the enlightenment thinkers pushed for more consideration of individual subjective experience, much of their considerations of beauty still gave priority to objective considerations of beauty, such as order and proportion.

It wasn't until the 18th century that subjective experience really became the centerpiece of aesthetic consideration. The 18th century birthed the romantic period, a reaction against the objectivism and scientism of the enlightenment. Specifically, the romantics focused on the reality of subjective experience through the individual's passion. This means that rather than order and reason, the romantics were concerned to elaborate the reality of an individual's life through their emotions, attitudes about the world, and desires.

So, there are a few questions we can ask about the subjective view of beauty. In this view, we take it for granted that all beautiful experiences are a quality of the subject's experience of a thing. The thing experienced (be it artwork, person, mountain, etc.) doesn't have any beauty in itself, instead it is only beautiful insofar as an individual ascribes beauty to it. For this reason, it is hard to tell what beauty is an experience of – personal taste, pleasure, or if it is simply an aspect of the entertainment value of the thing. It also makes it difficult to determine whether or not that thing is really beautiful: if I can't stand *Game of Thrones* but my friends love it, who is correct? If beauty is only in the eye of the beholder, then are we both correct, and *Game of Thrones* is both ugly and beautiful? Without an independent standard it's difficult to tell if someone else is wrong, or if beauty is even worth talking about at all.

One version of the subjective argument is that all individual attributions of beauty have to do with the individual's taste. Taste is an attitude we have toward our experiences which informs us feeling pleasure or displeasure. This means that our individual taste informs whether we will feel pleasure when looking at something. If we find that experiencing the object is pleasurable, then we say that it is beautiful. This view, of course, shares a lot with social constructionism. **Social constructionism** is the view that all values exist simply because individuals in a society have agreed upon an arbitrary value and elevated it above others based on that agreement. "Beauty" is just a convention, or a way of referring to an object that gives us pleasure, which is the important part of the experience.

Our judgments about taste are also something inherently communicable. When we say "Oh, I like that!" about a painting at an art gallery, we're not merely informing ourselves. We want to share our taste, and to discuss the merits or demerits of the thing we're experiencing. Similarly, if beauty is only a convention for referring to things we find pleasurable, this convention must come from somewhere. Often, this somewhere is our upbringing or education, and this also has to do with our ability to communicate the things we find pleasurable with other people who are similarly predisposed to us (whether of the same class, or identity, and so on).

So, beauty can either be a quality independent of the thing, or it can be something we subjectively experience about the thing. In the first case, as an independent objective quality, beauty must make up part of the existence of the thing. The beautiful thing is really beautiful because we are experiencing something about that thing itself. In the second case, as a subjective quality of our experience, beauty isn't a real aspect of the thing at all. Instead, beauty is only part of the appearance of the thing; in fact, what it is in itself doesn't matter at all! If beauty is only part of the appearance of the thing, then it's just a property of our subjective experiences, such as a particular feeling we have about it or a reaction we have to it. Since we like the artwork, it must be beautiful.

In this section we have discussed the two major views on what a beautiful thing is: whether beauty is an objective quality of the thing, or whether beauty is a subjective quality of our experience of the thing. For the objective view, aspects of the form of the thing often matter most, such as order and proportion. For the subjective view, aspects of the content of the thing often matter more. The content of an artwork is often what we attach emotions to, such as beautiful scenery or a tragic scene. The form can also influence our subjective

attitudes toward the artwork, though this often depends more on the artwork than our individual attitude. For the subjectivist, how I identify with the thing and want to communicate that to others (especially others that see it the way I do) is what makes up the beauty of it. These two views are not mutually exclusive, and in the history of philosophy many thinkers have grabbed a little from the objective view and a little from the subjective view. However, this distinction is helpful for understanding why the existence of beauty and the distinction between form and content are so important for other considerations about art specifically. We will now move on to discuss aesthetic experience as it pertains to art specifically, rather than the experience of beauty generally.

Study questions

1. True or False: The romantics were reacting against the enlightenment philosophers' idea of beauty.
2. True or False: The Ancient and Classical philosophers share similar views about beauty.
3. True or False: Beauty is only subjective.
4. True or False: Formalists are objectivists about beauty.
5. True or False: Judgments of taste are central to subjectivists' account of beauty.
6. True or False: The artwork's subject matter or meaning is called its form.
7. True or False: The artwork's subject matter or meaning is called its content.
8. True or False: A social constructionist would only think Van Gogh's paintings are beautiful if the rest of society does.
9. True or False: Romantic theories about art can be called subjectivist, due to their reliance on individual feeling produced by an artwork.
10. True or False: Enlightenment theories about art can be called formalist, since they rely heavily on the technical, proportional, and mathematical perfection of an artwork.

For deeper thought

1. Can someone be an objectivist about beauty, but not about art? Is the value of beauty distinct from the value of an artwork?
2. Can people desire non-beautiful things according to the definition of beauty given above?
3. Can we judge others' taste if we are objectivists about art – ie, can we say someone has bad taste if they dislike a piece of artwork we think is beautiful because of its formal qualities (like the Darmstadt Madonna)?

What are works of art?

Often, when we think of beauty, we think of great artworks. Though there are other forms of beauty, such as mathematical or natural, the most prominent notion of beauty is that given in art. In this section we will cover some philosophical discussions over what art is, the relationship between artworks and other objects, the value and function of art, and the nature of representation. By the end of this section, you should be able to apply the two views of beauty to specific kinds of art.

First, to define art, we should distinguish it from other kinds of beauty. Artworks are non-natural, since they are made by humans. In terms of the objects we experience, we can broadly say there are two kinds. **Natural objects** have their purpose internal to them – an acorn, for instance, exists for the sake of becoming a tree. **Artifacts** have a purpose external to them, since they are made by something that has a plan for them. Chairs, for instance, are made by humans to sit on. You could eat on a chair, even though that isn't it's given purpose. However, most people would find you odd, since humans build chairs for sitting. In this way, there is a very basic dichotomy between natural objects and artifacts.

Artworks are a form of artifact. **Artworks** can generally be defined as any artifact which expresses some kind of idea. Additionally, artworks often function primarily as objects of beauty and contemplation, rather than instruments. The idea doesn't need to be intentional – a vase from ancient Greece might tell us much about the values of ancient Grecians. It also doesn't need to be abstract or intellectual – Monet's paintings could simply express his technique for capturing living scenery. Artworks will never be natural objects, either. When we experience natural beauty, we don't wonder about the technique used to create it nor the ideas expressed in the object. Artwork is also not the sole experience of beauty, though it has a special purpose in helping us understand our place in the world that other experiences of beauty lack.

Our broad definition of artworks does seem to have some difficulties to it, though. Surely there must be some difference between one of those Monet landscapes and my doodles to show my landlord where the water pipe broke! Or, as a more plausible example, there should be some kind of distinction between artworks like comics that are fairly straightforward, and artworks like a Monet which may express much more than a Garfield comic. In the 18th century, the romantics came up with such a distinction (largely due to their obsession with cultivating taste): high art and low art. **High art** is usually intellectual and takes refined taste to appreciate. **Low art,** by contrast, is art merely for the sake of pleasure. While Monet may be high art, Garfield is certainly low art in this distinction. This is certainly not a dichotomy that has been accepted universally, nor is it really a good description of the difference between Monet and Garfield. However, it is a helpful interpretive device to start thinking about the differences between them.

This distinction gets especially dicey with ancient art. Greek vases, for instance, are central pieces of classical art. The vases often have depictions of gods in some act or illustrate some aspect of Greek mythology. These

were simply vases to fill with wine, or foodstuff, or anything else. Yet these vases tell us so much about Greek life in that period that they are regarded with the same interest as classical Greek sculpture (which fits much easier into "high art"). They are also done with great craftsmanship, but the same claim could be made about hand-me-down blankets that our grandmothers made in the great depression. While it seems like the high/low art dichotomy has to do with the intention of the artwork, it also has to do with the subject matter of the art.

We have all found ourselves at some point at an art museum, looking at some weird cobbled together sculpture, wondering "What could this possibly mean?" Many of us approach art in this way, expecting a communicable meaning somewhere within the artwork. This is especially true since the advent of conceptual art in the last 50 years, which often focuses on conveying some kind of message (usually moral). While this is a dominant manner of interpreting art, it is not the only idea about meaning in art, and in the last few centuries other ideas have become dominant. The idea that art needs to stand for something else is only one possible interpretation of what is presented in an artwork. The three possibilities we will cover here are representation, expression, and form.

Theories of meaning in art

Historically, the dominant theory of meaning in art is the representational theory. **Representation** means that an something stands for another thing. An image or symbol can represent an abstract idea or simply represent the subject it portrays. The represented meaning is inherently communicable and an objective fact about the artwork. Since Plato's time, art has been predominantly considered a representational activity. **The representation theory of art** is that the meaning of an artwork comes from what it represents. The representation can simply be some thing that is painted, like in a landscape painting. The representation can also use the straightforward image in the painting (say, the landscape painting) to represent an abstract idea, which is not immediately implied by the image in the painting. The landscape could represent the wildness of nature, or the need for us Americans to protect our natural resources. It could also just simply represent that landscape like a photograph would, by giving us an idea of what it is like to be in that specific natural setting.

You can see in this photorealist painting by John Bader, Johns' Diner with John's Chevelle, a level of perfection in realism that you do not see in other forms of landscape painting like Van Gogh's painting above. Believe it or not, this is a painting, not a photograph! If we were concerned simply with the realistic portrayal of a landscape, it would make Bader's painting objectively more beautiful. Though there are certainly other arguments to make about whether one painting is better or not, for someone who is only interested in the correct representation of nature through art, Bader's painting is objectively better.

The romantics, however, felt the representational theory of art was insufficient. For the romantics, nothing about the communicable and objective meaning of an artwork explains why we have the taste for some forms of art and why some art forms are so moving. The romantics instead thought that the meaning of art was in its expression. **Expression** in this sense means that the subject of the artwork is the emotion attached to the image. **The romantic theory of art** is that art does not convey a concept but a powerful emotion or feeling. A good example of this is the romantic painting *Wanderer Above the Sea of Fog* by Caspar David Friedrich. While the painting may not evoke a specific emotion, such as sadness, we do not need to 'read' the painting in a specific way to understand its objective meaning. Instead, the painting evokes some kind of strong feeling: that of solitude, or overcoming, or perhaps wonder at the magnitude of nature.

Caspar David Friedrich, Wanderer Above the Sea of Fog

Even after the romantics, many forms of art defy either representational or romantic theories of art. This is especially true of modern and postmodern art. Modern art specifically defines the meaning of art by its form. Form, remember, is the composition or technique used to create the artwork. In many of these art forms the point of the artwork is to allow the spectator to reflect on what an artwork does and how their perception is formed by different techniques of constructing the world. This is especially prevalent in Picasso's cubism, where the painting is meant to show we understand the objectivity of an object from the different possible viewpoints we can have on it. This kind of art is more of a 'meta' art, so we can call it reflective art. The **reflective theory of art** requires that the content of a painting be a reflection on its form or presentation to the viewer.

The value of art

We've talked a lot about how to judge beauty and what beautiful art is. What about putting those conceptions of beauty to work and helping us make judgments of taste about the value of art – can we finally tell our friends our favorite TV show is better, and express why? In many current Hollywood movies, we assume that our identity and morals must be expressed. We assume the standard for others liking what we like must refer to a kind of social standard about what we believe. Many of these films – think about superhero movies like *Black Panther* and *Captain Marvel* – seem to also be merely produced for the purpose of pleasure. So, is it just that I like this thing or is it just that it is a socially useful tool for thinking morally? Neither of these options really captures why we like these movies and shows, though. This brings up an important difference in ideas about artistic value: should art have some external use or simply exist for its own sake?

This question doesn't just split according to the easy dichotomies we have previously discussed (though it can). Some objectivists like Plato think art should offer moral instruction, and some objectivists like the classical art theorists think art should simply portray perfect proportion and mastery of technique. Similarly, many subjectivists think that art must provide instruction for moral life, but just as many subjectivists think art is simply for pleasure. Often, in our culture, our view is a blend between the last two ideas: that beauty is subjective and should sometimes provide moral instruction, while sometimes it should just be for pleasure. Unfortunately, this doesn't help us decide on the value of art. How can I say that *CSI: Miami* is a better show, or even of the same artistic merit, as *Deadwood*? Moreover, can I compare *CSI: Miami* to Caspar David Friendrich's *Wanderer Above the Sea of Fog*? These comparisons will seriously depend on your view of beauty, and thus your views on art. If you think art is only subjectively beautiful, you might appeal to others' taste and argue that what *CSI: Miami* expresses is more interesting than *Wanderer Above the Sea of Fog*. You could also be an objectivist and claim that all of the TV shows are bad examples of art, since none represent mathematical or technical perfection in their portrayal of time using the montage like the old black and white film *Battleship Potemkin*, and thus do not portray beauty through their representations. These are both pretty extreme examples, but they show that to make coherent arguments about art, we need to consider:

1. What the quality is in the artwork that draws us to it (beauty – objective or subjective)
2. What differentiates it from other artworks (taste – objective or subjective)
3. What makes it art rather than not (representation, expression, reflection)

Value also depends, at least in some part, on the intention of the artist. Is the value of art determined by the intention of the artist, or is it simply the effect of the artwork itself? We might think back to those Greek sculptures – are those great pieces of art because some Greek sculptor thought to himself "I must express the form of beauty to instruct the masses to live a moral life!" or are they simply great pieces of art because they show us what it meant to be a Greek? If we look back on many of the great pieces of art in history, many of them are simply great because they advance the techniques of their art so far or because they capture the spirit of their age so well. None of these works, however, is great because the idea in the author's head is so masterfully represented in the work. Rather, it is because the important sociocultural impact of these art pieces that they are remembered.

Issues about the value and intention behind the work of art are just a few of many current problems within aesthetic theory. With the resurgence of traditional forms of classical art in music and painting, the reflective content of film and video games, and the relationship between pleasure and morality in mass culture, it is difficult to say there are any unified views on art anymore that can account for all contemporary artistic production. Though some of the classical views hold sway, so do the romantic and postmodernist views. We will end this chapter with some important contemporary questions to consider about aesthetic theory, beauty, and the nature of art posed in the question "For deeper thought" section, below.

Study questions

1. True or False: Natural objects have an internal purpose, whereas artifacts must be given a purpose by something else.
2. True or False: Artworks are artifacts which serve as objects of beauty and contemplation.
3. True or False: Expressive art refers to art that makes you feel some kind of emotion about the content portrayed in it.
4. True or False: Representational art refers to art that portrays an idea or symbol.
5. True or False: Reflective art is art that takes questions about its own meaning as its content via reflection on its form.
6. True or False: Natural objects can have beauty.
7. True or False: Natural objects can be art.
8. True or False: According to theories which accept the dichotomy, high art is more intellectual than low art.
9. True or False: According to theories which accept the dichotomy, comic books are low art.

10. True or False: The Darmstadt Madonna pictured above is considered beautiful thanks to its formal perfection, rather than its emotional meaning.

For deeper thought

1. Should good art be pleasurable? Is 'being entertained' the same as being pleased?
2. Should good art communicate a message, moral or otherwise? Does art need to reflect upon the form of its presentation?
3. Are there forms of art that have meaning to us that do not rely on our judging them beautiful, even if this beauty does not refer to objective perfection?
4. If we judge contemporary art meaningful even if it is not pleasurable, does this mean that we have changed our definition of beauty or does it mean that not all art is beautiful?

INDEX

B

C

D

E

F

K

L

Q

R

S

T

U

V

Valid: 20, 116, 122

Valid argument: 20

W

Well-being: 156-167, 169, 170, 171, 218

Working class: 204