A Brief Introduction to Philosophy

A Brief Introduction to Philosophy

YONI PORAT

SOUTHERN ALBERTA INSTITUTE OF TECHNOLOGY
CALGARY, ALBERTA, CANADA

Contents

Part VIII. Introducing Moral Theories

Preface

Great News! Your course is using an Open Education digital textbook!

What does Open Education mean?

Open Education textbooks are a type of Open Education Resource (OER) that have been licensed to be freely used, adapted, and distributed. **Open Education textbooks are free for you to use.** This book is licensed with a Creative Commons Attribution, Non-Commercial, Share-Alike 4.0 license (CC BY NC SA 4.0) **which allows you to make changes, save, and print this book as long as the use is non-commercial. Unlike other ebooks, this book does not require a username and password, and you can keep it forever!**

What does digital mean?

This textbook is provided in multiple file types that you can access on most smartphones, tablets, and computers. The web-based format means additional information, such as alternative image tags

and live links, can be included. You can also download a PDF file that has been optimized to work with the most common forms of adaptive technology, or a file type that works with an e-book reader. **You can use handy features such as highlighting, note-taking, and using *ctrl-f* to find concepts quickly within the e-book to help you with your studies.** Copy and paste sections into your favourite flashcard or study system and share with friends!

Can I print a copy of this book?

Yes, the open license means that you can print any or all parts of this book for personal use. For example, you can take the PDF file to the X-Docs Centre on campus and have the book printed in black and white, double sided, coil bound for approximately $25.00 + GST. (Costs may change without notice.)

(OER logo by Skyline College is licensed under CC BY 4.0)

About the Author

The author of this textbook deserves little credit in comparison with the original authors from whom he excerpted this material (see the attributions sections for the specifics) and the extraordinary librarian who undertook the mechanics of coordinating the process. The author will accept credit for being an enthusiastic educator who loves teaching philosophy and guiding students to being more philosophical. He is intensely sympathetic to the vision of Open Education, so that expense will be less of an obstacle to participation in formal education.

Yoni Porat has been teaching for over twenty years, ten of those focusing on Philosophy.

Licensing Information

This work is made available in 2021 by Southern Alberta Institute of Technology (SAIT) under the terms of a Creative Commons Attribution-NonCommercial-ShareAlike licence, except where otherwise noted.

The cover image is adapted from "Rodin Thinker inspired in CDs" by Rudy and Peter Skitterians (2017), available under a CC0 (Public Domain) Licence. Attribution for resources used within the work are found at the end of each chapter. The SAIT name and SAIT logo are not included within this licence and may not be reproduced without the prior and express written permission of Southern Alberta Institute of Technology.

Acknowledgements

This project is supported by the Reg Erhardt Library at SAIT, in partnership with University of Alberta Library's Open Education Alberta publishing project.

Supported by

In partnership with

Special thanks go to folks at SAIT who provided expertise: Brian Dinnigan, Curriculum Assistant, and Jessica Norman, OER Librarian, for their work in the areas of ebook publishing and open licensing. We also thank Haley Kleckner, SAIT LIT Practicum Student, for her countless hours of clerical and editing support.

If you adopt this book, as a core or supplemental resource, please report your adoption in order for us to celebrate your support of students' savings. We also invite you to adapt this book further to meet your and your students' needs. Please let us know if you do so by sending an email to library@sait.ca.

PART I

PHILOSOPHY FUNDAMENTALS

1. Introduction

If this is your first exposure to the study of philosophy then you will quickly notice that philosophy is both seriously weird and weirdly serious. It is seriously weird because it takes on topics that don't fit into other areas of study. This isn't a coincidence; philosophy has a long history of asking questions before there's enough clarity to start answering those questions. Once there is enough clarity then a new type of knowledge is born. Before science there was natural philosophy. Before psychology there was philosophy of mind. Before an area becomes systematic and comprehensible, philosophy explores the landscape, mapping at least the more interesting properties and what connections there seem to be. So, almost by definition, if a question is philosophical it will likely seem strange.

Philosophy is also weirdly serious since it doesn't brush past questions which are often otherwise dismissed. How can you have free will if you are just a biological machine? Why would an all-powerful, benevolent, and wise deity create evil? What does it mean for something to exist?

You can imagine asking these questions without expecting an answer; in this book you will read what philosophers who took those questions seriously came up with. Whether or not you agree with those answers, you will see that it is possible to make progress, to not just give up without any answer. If those questions and the other ones you'll see in this book matter to you, then being able to answer them will be a matter of the utmost seriousness for you.

This introduction to philosophy will touch on some key questions that have been around for centuries, that neither science nor religion have addressed to the general satisfaction of those who wonder. This book isn't comprehensive; there is more to be said on every one of these issues. This book isn't authoritative; there are other voices to be heard and approaches to attempt to address the same questions. This book isn't conclusive; if you find that there's

more that you need to know then this will be just the beginning for you. This book is a starting point. Hopefully, it will also be a source of inspiration for your philosophical journey.

-Yoni Porat (Calgary, August 2021)

2. What Philosophy Is

What is philosophy?

Many answers have been offered in reply to this question and most are angling at something similar. One answer is that philosophy is that it is the love of wisdom (Philo = love, Sophos = Wisdom).

Perhaps you think science exhausts inquiry. About a hundred years ago, many philosophers, especially the Logical Positivists, thought there was nothing we could intelligibly inquire into except for scientific matters. But this view is probably not right. What branch of science addresses the question of whether or not science covers all of rational inquiry? If the question strikes you as puzzling, this might be because you already recognize that whether or not science can answer every question is not itself a scientific issue. Questions about the limits of human inquiry and knowledge are philosophical questions.

We can get a better understanding of philosophy by considering what sorts of things other than scientific issues humans might inquire into. Philosophical issues are as diverse and far ranging as those we find in the sciences, but a great many of them fall into one of four big topic areas: metaphysics, epistemology, logic, and ethics.

Metaphysics

Metaphysical issues are concerned with the nature of reality. Traditional metaphysical issues include the existence of God and the nature of human free will (assuming we have any). Here are

a few metaphysical questions of interest to contemporary philosophers: What is it that makes a thing a thing? How are space and time related? Does the past exist? How about the future? How many dimensions does the world have? Are there any entities beyond physical objects (like numbers, properties, and relations)? If so, how are they related to physical objects? Historically, many philosophers have proposed and defended specific metaphysical positions, often as part of systematic and comprehensive metaphysical views. But attempts to establish systematic metaphysical world views have been notoriously unsuccessful.

Since the 19th century many philosophers and scientists have been understandably suspicious of metaphysics, and it has frequently been dismissed as a waste of time, or worse, as meaningless. But in just the past few decades metaphysics has returned to vitality. As difficult as they are to resolve, metaphysical issues are also difficult to ignore for long. Contemporary analytic metaphysics is typically taken to have more modest aims than definitively settling on the final and complete truth about the underlying nature of reality. A better way to understand metaphysics as it is currently practiced is as aiming at better understanding how various claims about reality logically hang together or conflict. Metaphysicians analyze metaphysical puzzles and problems with the goal of better understanding how things could or could not be. Metaphysicians are in the business of exploring the realm of possibility and necessity. They are explorers of conceptual space.

Epistemology

Epistemology is concerned with the nature of knowledge and justified belief. What is knowledge? Can we have any knowledge

at all? Can we have knowledge about the laws of nature, the laws or morality, or the existence of other minds? The view that we can't have knowledge is called skepticism. An extreme form of skepticism denies that we can have any knowledge whatsoever. But we might grant that we can have knowledge about some things and remain skeptics concerning other issues. Many people, for instance, are not skeptics about scientific knowledge, but are skeptics when it comes to knowledge of morality. Some critical attention reveals that scientific knowledge and moral knowledge face many of the same skeptical challenges and share some similar resources in addressing those challenges. Many of the popular reasons for being more skeptical about morality than science turn on philosophical confusions.

Even if we lack absolute and certain knowledge of many things, our beliefs about those things might yet be more or less reasonable or more or less likely to be true given the limited evidence we have. Epistemology is also concerned with what it is for a belief to be rationally justified.

Even if we can't have certain knowledge of anything (or much), questions about what we ought to believe remain relevant.

Logic

Logic is the study of arguments. Informal logic involves looking at different types of arguments and distinguishing the good from the bad. Formal logic looks at arguments solely in terms of their form. It is a type of mathematics on language, where the only values are true or false, 0 or 1. It can be used to evaluate whether the argument is logically perfect and also what conclusions can be drawn from it with absolute confidence.

Ethics

While epistemology is concerned with what we ought to believe and how we ought to reason, Ethics is concerned with what we ought to do, how we ought to live, and how we ought to organize our communities. Sadly, it comes as a surprise to many new philosophy students that you can reason about such things. Religiously inspired views about morality often take right and wrong to be simply a matter of what is commanded by a divine being or beings. Cultural Relativism, perhaps the most popular opinion among people who have rejected faith, simply substitutes the commands of society for the commands of God or gods. Commands are simply to be obeyed, they are not to be inquired into, assessed for reasonableness, or tested against the evidence. Thinking of morality in terms of whose commands are authoritative leaves no room for rational inquiry into how we ought to live, how we ought to treat others, or how we ought to structure our communities. Philosophy, on the other hand, takes seriously the possibility of rational inquiry into these matters. If philosophy has not succeeded in coming up with absolutely certain and definitive answer in ethics, this is in part because philosophers take the answers to moral questions to be things we need to discover, not simply matters of somebody's say so. The long and difficult history of science should give us some humble recognition of how difficult and frustrating careful inquiry and investigation can be. So we don't know for certain what the laws of morality are. We also don't have a unified field theory in physics. Why expect morality to be any easier?

We might think of metaphysics as concerned with "What is it?" questions, epistemology as concerned with "How do we know?" questions, and ethics as concerned with "What should we do about it?" questions. Many interesting lines of inquiry cut across these four kinds of questions. The philosophy of science, for instance, is concerned with metaphysical issues about what

science is, but also with both epistemological questions about how we can know scientific truths and logical questions about what has to be true. The philosophy of love is similarly concerned with metaphysical questions about what love is.

But it also concerned with questions about the value of love that are more ethical in character.

Assorted tangled vines of inquiry branch off from the four major trunks of philosophy, intermingle between them, and ultimately with other types of knowledge as well. The notion that some branches of human inquiry can proceed entirely independent of others ultimately becomes difficult to sustain.

What is the value of philosophy?

Philosophy is a branch of human inquiry and as such it aims at knowledge and understanding. We might expect that the value of philosophy lies in the value of the ends that it seeks, the knowledge and understanding it reveals. But philosophy is rather notorious for failing to establish definitive knowledge on the matters it investigates. I'm not so sure this reputation is well deserved. We do learn much from doing philosophy. Philosophy often clearly reveals why some initially attractive answers to big philosophical questions are deeply problematic, for instance. Granted, philosophy often frustrates our craving for straightforward convictions. In our first reading, Bertrand Russell argues that there is great value in doing philosophy precisely because it frustrates our desire for quick and easy answers. In denying us easy answers to big questions and undermining complacent convictions, philosophy liberates us from narrow minded conventional thinking and opens our minds to new possibilities. Philosophy often provides an antidote to prejudice not by settling big questions, but by revealing just how

hard it is to settle those questions. It can lead us to question our comfortably complacent conventional opinions.

3. The Value of Philosophy

The Value of Philosophy

Reading: The first Reading is Chapter 15 of Bertrand Russell's *Problems of Philosophy*, "The Value of Philosophy." The whole book can be found here:
 http://www.ditext.com/russell/russell.html. (Follow one of these links and do the reading. Then, read the discussion of it below)

We humans are very prone to suffer from a psychological predicament we might call "the security blanket paradox." We know the world is full of hazards, and like passengers after a shipwreck, we tend to latch on to something for a sense of safety. We might cling to a possession, another person, our cherished beliefs, or any combination of these. The pragmatist philosopher Charles Sanders Peirce speaks of doubt and uncertainty as uncomfortable anxiety-producing states. This would help explain why we tend to cling, even desperately, to beliefs we find comforting. This clinging strategy, however, leads us into a predicament that becomes clear once we notice that having a security blanket just gives us one more thing to worry about. In addition to worrying about our own safety, we now are anxious about our security blanket getting lost or damaged. The asset becomes a liability. The clinging strategy for dealing with uncertainty and fear becomes counterproductive.

While not calling it by this name, Russell describes the intellectual consequences of the security blanket paradox vividly:

> The man who has no tincture of philosophy goes through life imprisoned in the prejudices derived from

common sense, from the habitual beliefs of his age or his nation, and from convictions which have grown up in his mind without the cooperation or consent of his deliberate reason. . . The life of the instinctive man is shut up within the circle of his private interests. . . In such a life there is something feverish and confined, in comparison with which the philosophic life is calm and free. The private world of instinctive interests is a small one, set in the midst of a great and powerful world which must, sooner or later, lay our private world in ruins.

The primary value of philosophy according to Russell is that it loosens the grip of uncritically held opinion and opens the mind to a liberating range of new possibilities to explore.

The value of philosophy is, in fact, to be sought largely in its very uncertainty. . . Philosophy, though unable to tell us with certainty what is the true answer to the doubts which it raises, is able to suggest many possibilities which enlarge our thoughts and free them from the tyranny of custom. Thus, while diminishing our feeling of certainty as to what things are, it greatly increases our knowledge as to what they may be; it removes the somewhat arrogant dogmatism of those who have never traveled into the region of liberating doubt, and it keeps alive our sense of wonder by showing familiar things in an unfamiliar aspect.

Here we are faced with a stark choice between the feeling of safety we might derive from clinging to opinions we are accustomed to and the liberation that comes with loosening our grip on these in order to explore new ideas. The paradox of the security blanket should make it clear what choice we should consider rational. Russell, of course, compellingly affirms choosing the liberty of free and open inquiry.

Must we remain forever uncertain about philosophical matters? Russell does hold that some philosophical questions

appear to be unanswerable (at least by us). But he doesn't say this about every philosophical issue. In fact, he gives credit to philosophical successes for the birth of various branches of the sciences. Many of the philosophical questions we care most deeply about, however – like whether our lives are significant, whether there is objective value that transcends our subjective interests – sometimes seem to be unsolvable and so remain perennial philosophical concerns. But we shouldn't be too certain about this either. Russell is hardly the final authority on what in philosophy is or isn't resolvable. Keep in mind that Russell was writing about a century ago and a lot has happened in philosophy in the mean time (not in small part thanks to Russell's own definitive contributions). Problems that looked unsolvable to the best experts a hundred years ago often look quite solvable by current experts. The sciences are no different in this regard. The structure of DNA would not have been considered knowable fairly recently. That there was such a structure to discover could not even have been conceivable prior to Mendel and Darwin (and here we are only talking less than two centuries ago).

Further, it is often possible to make real progress in understanding issues even when they can't be definitively settled. We can often rule out many potential answers to philosophical questions even when we can't narrow things down to a single correct answer. And we can learn a great deal about the implications of and challenges for the possible answers that remain.

Even where philosophy can't settle an issue, it's not quite correct to conclude that there is no right answer. When we can't settle an issue this usually just tells us something about our own limitations. There may still be a specific right answer; we just can't tell conclusively what it is. It's easy to appreciate this point with a non-philosophical issue. Perhaps we can't know whether or not there is intelligent life on other planets. But, surely either there is or there isn't intelligent life on other

planets. Similarly, we may never establish that humans do or don't have free will, but it still seems that there must be some fact of the matter. It would be intellectually arrogant of us to think that a question has no right answer just because we aren't able to figure out what that answer is.

Review and Discussion Questions

The following questions will help you organize your thoughts about the past few chapters. Feel free to take these questions up on the discussion board.

On this lecture note:

- Why should we doubt that science covers all of human inquiry?

- What are some metaphysical issues? Some epistemological and ethical issues?

-
- What problem does the view that morality is simply a matter of the say-so of some authority lead to?

On Russell's "The Value of Philosophy":

   ◦ What is the aim of philosophy according to Russell?
   ◦ How is philosophy connected to the sciences?
   ◦ What value is there in the uncertainty that philosophical inquiry often produces?

In Chapter 3, the commentary on Russell:

- Explain the "security blanket" paradox.
- How can understanding of issues be advanced even when definitive knowledge can't be had?
- What's the difference between saying we can't know the

answer to some question and saying that there is no truth of the matter?

Finally, consider some of the definitions of philosophy offered by philosophers on: http://www.brainpickings.org/index.php/2012/04/09/what-is-philosophy/
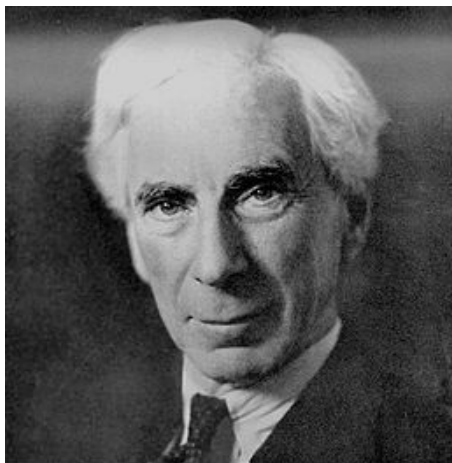
4. Bertrand Russell

Bertrand Arthur William Russell, 3rd Earl Russell, 1872 – 1970 CE, was a British philosopher, writer, social critic and political activist. In the early 20th century, Russell led the British "revolt against idealism." He is considered one of the founders of analytic philosophy. Russell was an anti-war activist and went to prison for his pacifism during World War I.   He did conclude that the war against Adolf Hitler was a necessary "lesser of two evils." He won the Nobel Prize in Literature in 1950 "in recognition of his varied and significant writings in which he champions humanitarian ideals and freedom of thought."

In "Reflections on My Eightieth Birthday" ("Postscript"

in his *Autobiography*), Russell wrote: "I have lived in the pursuit of a vision, both personal and social.

> **Personal:** to care for what is noble, for what is beautiful, for what is gentle; to allow moments of insight to give wisdom at more mundane times.

> **Social:** to see in imagination the society that is to be created, where individuals grow freely, and where hate and greed and envy die because there is nothing to nourish them. These things I believe, and the world, for all its horrors, has left me unshaken".

You might find it interesting to see the two things that he believed he would like to say to a future generation. It takes less than 2 minutes to watch; in 1959, this is what Bertrand Russell had to say (click on the link):

<div align="center">Message to Future Generations</div>

5. Plato's "Allegory of the Cave"

Allegory of the Cave

Socrates: And now, I said, let me show in a figure how far our nature is enlightened or unenlightened:

Behold! human beings living in a underground den, which has a mouth open towards the light and reaching all along the den; here they have been from their childhood, and have their legs and necks chained so that they cannot move, and can only see before them, being prevented by the chains from turning round their heads. Above and behind them a fire is blazing at a distance, and between the fire and the prisoners there is a raised way; and you will see, if you look, a low wall built along the way, like the screen which marionette players have in front of them, over which they show the puppets.

Glaucon: I see.

Socrates: The low wall, and the moving figures of which the shadows are seen on the opposite wall of the den. And do you see, I said, men passing along the wall carrying all sorts of vessels, and statues and figures of animals made of wood and stone and various materials, which appear over the wall? Some of them are talking, others silent.

Glaucon: You have shown me a strange image, and they are strange prisoners.

Socrates: Like ourselves, I replied; and they see only their own shadows, or the shadows of one another, which the fire throws on the opposite wall of the cave?

Glaucon: True, how could they see anything but the shadows if they were never allowed to move their heads?

Socrates: And of the objects which are being carried in like manner they would only see the shadows?

Glaucon: Yes.

Socrates: And if they were able to converse with one another, would they not suppose that they were naming what was actually before them

Glaucon: Very true.

Socrates: The prisoners would mistake the shadows for realities. And suppose further that the prison had an echo which came from the other side, would they not be sure to fancy when one of the passers-by spoke that the voice which they heard came from the passing shadow?

Glaucus: No question.

Socrates: To them, I said, the truth would be literally nothing but the shadows of the images.

Glaucon: That is certain.

Socrates: And now look again, and see what will naturally follow if the prisoners are released and disabused of their error. At first, when any of them is liberated and compelled suddenly to stand up and turn his neck round and walk and look towards the light, he will suffer sharp pains; the glare will distress him, and he will be unable to see the realities of

which in his former state he had seen the shadows; and then conceive some one saying to him, that what he saw before was an
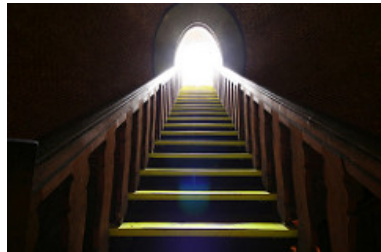
illusion, but that now, when he is approaching nearer to being and his eye is turned towards more real existence, he has a clearer vision,—what will be his reply? And when released, they would still persist in maintaining the superior truth of the shadows. And you may further imagine that his instructor is pointing to the objects as they pass and requiring him to name them,—will he not be perplexed? Will he not fancy that the shadows which he formerly saw are truer than the objects which are now shown to him?

Glaucon: Far truer.

Socrates: And if he is compelled to look straight at the light, will he not have a pain in his eyes which will make him turn away to take refuge in the objects of vision which he can see, and which he will conceive to be in reality clearer than the things which are now being shown to him?

Glaucon: True.

Socrates: When dragged upwards, they would be dazzled by excess of light. And suppose once more, that he is reluctantly dragged up a steep and rugged ascent, and held fast until he is forced into the presence of the sun himself, is he not likely to be pained and irritated? When he approaches the light his eyes will be dazzled, and he will not be able to see anything at all of what are now called realities.

Glaucon: Not all in a moment.

Socrates: He will require to grow accustomed to the sight of the upper world. And first he will see the shadows best, next the reflections of men and other objects in the water, and then the objects themselves; then he will gaze upon the light of the moon and the stars and the spangled heaven; and he will see the sky and the stars by night better than the sun or the light of the sun by day?

Glaucon: Certainly.

Socrates: Last of all he will be able to see the sun, and not mere reflections of him in the water, but he will see him in his own proper place, and not in another; and he will contemplate him as he is.

Glaucon: Certainly.

Socrates: He will then proceed to argue that this is he who gives the season and the years, and is the guardian of all that is in the visible world, and in a certain way the cause of all things which he and his fellows have been accustomed to behold?

Glaucon: Clearly, he would first see the sun and then reason about him.

Socrates: They would then pity their old companions of the den. And when he remembered his old habitation, and the wisdom of the den and his fellow-prisoners, do you not suppose that he would felicitate himself on the change, and pity them?

Glaucon: Certainly.

Socrates: And if they were in the habit of conferring honors among themselves on those who were quickest to observe the passing shadows and to remark which of them went before, and which followed after, and which were together; and who were therefore best able to draw conclusions as to the future, do you think that he would care for such honors and glories, or envy the possessors of them? Would he not say with Homer,

'Better to be the poor servant of a poor master,'

and to endure anything, rather than think as they do and live after their manner?

Glaucon: Yes. I think that he would rather suffer anything than entertain these false notions and live in this miserable manner.

Socrates: Imagine once more, I said, such an one coming suddenly out of the sun to be replaced in his old situation; would he not be certain to have his eyes full of darkness?

Glaucon: To be sure.

Socrates: And if there were a contest, and he had to compete in measuring the shadows with the prisoners who had never moved out of the den, while his sight was still weak, and before his eyes had become steady (and the time which would be needed to acquire this new habit of sight might be very considerable), would he not be ridiculous? Men would say of him that up he went and down he came without his eyes; and that it was better not even to think of ascending; and if any one tried to loose another and lead him up to the light, let them only catch the offender, and they would put him to death.

Glaucon: No question.
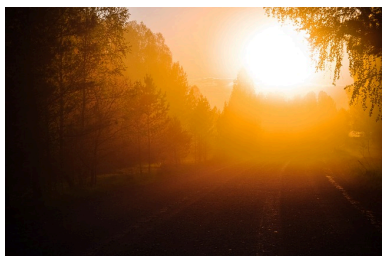
Socrates: The prison is the world of sight, the light of the fire is the sun. This entire allegory, I said, you may now append, dear Glaucon, to the previous argument; the prison-house is the world of sight, the light of the fire is the sun, and you will not misapprehend me if you interpret the journey upwards to be the ascent of the soul into the intellectual world according to my poor belief, which, at your desire, I have expressed—whether rightly or wrongly God knows. But, whether true or false, my opinion is that in the world of knowledge the idea of good appears last of all, and is seen only with an effort; and, when seen, is also inferred to be the universal author of all things beautiful and right, parent of light and of the lord of light in this visible world, and the immediate source of reason and truth in the intellectual; and that this is the power upon which he who would act rationally either in public or private life must have his eye fixed.

Glaucon: I agree, as far as I am able to understand you.

6. How Philosophy is Done

How Philosophy is Done

As a kind of inquiry, philosophy is aimed at establishing knowledge and understanding. Even where certain knowledge about a particular issue can't be had, there are often interesting things to learn about why we can't have certainty and what sorts of less-than-certain reasons there are for or against holding a position on that issue. So, rational inquiry may be interesting and fruitful even when we are denied straight-forward answers to our initial questions. Once we raise a philosophical issue, whether about the nature of justice or about the nature of reality, we want to ask what can be said for or against the various possible answers to our question. Here we are engaged in formulating arguments. Some arguments give us better reasons or accepting their conclusions than others. Once we have formulated an argument, we want to evaluate the reasoning it offers. If you want to know what philosophers do, this is a pretty good answer: philosophers formulate and evaluate arguments.

Your introduction to philosophy should be as much a training in how to do philosophy as it is a chance to get to acquainted with the views of various philosophers. To that end, you should carefully study the sections below on arguments.

Once a philosophical position is considered, we want to ask what arguments can be advanced in support of or against that issue. We then want to examine the quality of the arguments.

Evaluating flawed arguments often points the way towards other arguments and the process of formulating, clarifying, and evaluating arguments continues. This method of question and

answer in which we recursively formulate, clarify, and evaluate arguments is known as dialectic.

Dialectic looks a lot like debate, but a big difference lies in the respective goals of the two activities. The goal of a debate is to win by persuading an audience that your position is right and your opponent's is wrong. Dialectic, on the other hand, is aimed at inquiry. The goal is to learn something new about the issue under discussion. Unlike debate, in dialectic your sharpest critic is your best friend. Critical evaluation of your argument brings new evidence and reasoning to light. The person you disagree with on a philosophical issue is often the person you stand to learn the most from (and this doesn't necessarily depend on which of you is closer to the truth of the matter).

Dialectic is sometimes referred to as the Socratic Method after the famous originator of this systematic style of inquiry. We will get introduced to some of Plato's dialogues chronicling the exploits of Socrates in the next chapter on Ancient Greek Philosophy. This will give you a good sense for how the Socratic Method works. Then watch for how the Socratic Method is deployed throughout the rest of the course.

7. From Ancient to Modern Philosophy

From Ancient to Modern Philosophy

About 2000 years elapse between the ancient Greek philosophy and the modern classical period.

The rise and fall of Rome follows the golden age of ancient Greece. Greek philosophical traditions undergo assorted transformations during this period, but Rome is not known for making significant original contributions to either philosophy or science. Intellectual progress requires a degree of liberty not so available in the Roman Empire. Additionally, the intellectual talent and energy available in ancient Rome would have been pretty fully occupied with the demands of expanding and sustaining political power and order. Rome had more use for engineers than scientists, and more use for bureaucrats than philosophers. Christianity becomes the dominate religion in Rome after emperor Constantine converts in the 4th century A.D., Also in the 4th century, the great Christian philosopher Augustine, under the influence of

Plato, formulates much of what will become orthodox Catholic doctrine. After a rather dissolute and free-wheeling youth, Augustine studies Plato and find's much to make Christianity reasonable in it. With the rise of the Catholic Church, learning and inquiry are pursued largely exclusively in the service of religion for well over a millennium. Philosophy in this period is often described as the handmaiden of theology. The relationship between philosophy and theology is perhaps a bit more ambiguous, though. As we've just noted in the case of Augustine, much ancient Greek philosophy gets infused into

Catholic orthodoxy. But at the same time, the new faith of Christianity spearheads an anti-intellectual movement in which libraries are destroyed and most ancient Greek thought is lost to the world forever.

Through the West's period of Catholic orthodoxy, most of what we know of Greek science and philosophy, most notably Aristotle's thought, survived in the Islamic world. What remains of the complete works of Aristotle covers subjects as far ranging as metaphysics, ethics, politics, rhetoric, physics, biology, and astronomy, and amounts to enough writing to fill 1500 pages in the fine print translation on my bookshelf. But even this consists largely of lecture notes and fragments. Most of his polished prose is lost forever.

The crusade were a series of conflicts between the Christian and Islamic world towards the end of the middle ages. This conflict between Christianity and Islam was also an occasion for cultural exchange, and the Crusades led to the re-introduction of Aristotle and other ancient Greek scholarship to the west. Aristotle's philosophy and science was too carefully reasoned, systematic, and subtle to be dismissed as pointless pagan speculation. Instead, Christian thinkers in the west set out to understand Aristotle and interpret him a manner that would cohere with Catholic doctrine. St. Thomas Aquinas is the most famous philosopher to engage in this work of Christianizing Aristotle. He found ways to harness Aristotle's metaphysical arguments in the cause of advocating the existence of a Christian God.

Aristotle's views about the natural world quickly come to be received as the established truth in the Christian world. Aristotle's physics, for instance becomes the standard scientific view about the natural world in Europe. Aristotle also wrote about the methods of science, and he was much more empirical than his teacher Plato. Aristotle thought the way to learn about the natural world was to make careful observations and infer general principles from these. For instance, as an early biologist,

Aristotle dissected hundreds of species of animals to learn about anatomy and physiology. The Scholastics who studied Aristotle obviously did not adopt the methods Aristotle recommended. But some other people did. Galileo, Leonardo da Vinci, and Copernicus were among the few brave souls to turn a critical eye to the natural world itself and, employing methods Aristotle would have approved of, began to challenge the views of Aristotle that the Scholastics had made a matter of doctrine. Thus begins the Scientific Revolution.

Where the Renaissance is the reawakening of the West to its ancient cultural and intellectual roots, the Scientific Revolution begins as a critical response to ancient thinking, and in large part that of Aristotle. This critical response was no quick refutation. Aristotle's physics might now strike us as quite naïve and simplistic, but that is only because every contemporary middle school student gets a thorough indoctrination in Newton's relatively recent way understanding of the physical world. The critical reaction to Aristotle that ignites the scientific revolution grew out of tradition of painstakingly close study of Aristotle. The scholastic interpreters of Aristotle were not just wrongheaded folks stuck on the ideas of the past. They were setting the stage for new discoveries that could not have happened without their work. Again, our best critics are the ones who understand us the best and the one's from whom we stand to learn the most. In the Scientific Revolution we see a beautiful example of Socratic dialectic operating at the level of traditions of scholarship.

Europe also experiences significant internal changes in the 16th century that pave the way for its intellectual reawakening. In response to assorted challenges to the authority of the Catholic Church and the decadence of 16th century Catholic churchmen, Martin Luther launches the Reformation. The primary tenet of the reformation was that faith concerns the individual's relation to God who is knowable directly through the Bible without the intermediary of the Catholic Church. The

Reformation and the many splintering branches of Protestant Christianity that it spawns undermines the dogmatic adherence to a specific belief system and opens the way

for more free and open inquiry. The undermining of Catholic orthodoxy brought on by the reformation combined with the rediscovery of ancient culture in the Renaissance jointly give rise to the Scientific Revolution and, what we often refer to as the Modern Classical period in philosophy. The reawakening of science and philosophy are arguably one and the same revolution. Developments in philosophy and science during this period are mutually informed, mutually influencing, and intermingled. Individuals including Newton, Leibniz, and Descartes are significant contributors to both science and philosophy.

Review and Discussion Questions:

- Explain Protagoras' epistemic relativism.
- How does Socrates oppose epistemic relativism?
- What is the Socratic Method?
- How does Socrates respond to Euthyphro's suggestion that the pious is what is loved by all the gods? How does his response point us towards a critique of Divine Command Theory? What is the problem with the view that what is pious is pious because it is loved by the gods?
- What are Plato's forms? Why does Plato take the forms to be the most real sorts of entities?
- What is temperance and why is it a virtue in Plato's view?
- How is Plato's vision of justice non-egalitarian and anti-democratic?
- How do Plato and Aristotle's views on form differ?
- What is the difference between essence and accident?
- What does it mean to say that Aristotle held a teleological view

of the world?
- Explain Aristotle's four causes as principles of explanation.
- What is the role of Aristotle's philosophy and science in leading to the scientific revolution?

8. Attribution

The material in this chapter is adapted from previously published content:

Payne, R.W. (2015). *An Introduction to Philosophy.* BCcampus. Section 1.5-1.10, 2.11, 3.55. https://open.umn.edu/opentextbooks/textbooks/598 CC BY NC.

Ondich, J. (2018). *Words of Wisdom: Intro to Philosophy.* Minnesota Libraries Publishing Project. Sections 1.30-35, 41.409. https://mlpp.pressbooks.pub/introphil/ CC BY NC SA.

THE BASICS OF ARGUMENTS

9. What is an Argument?

What is an Argument?

This is an introductory textbook in logic and critical thinking. Both logic and critical thinking centrally involve the analysis and assessment of arguments. "Argument" is a word that has multiple distinct meanings, so it is important to be clear from the start about the sense of the word that is relevant to the study of logic. In one sense of the word, an argument is a heated exchange of differing views as in the following:

> Sally: Abortion is morally wrong and those who think otherwise are seeking to justify murder!
>
> Bob: Abortion is not morally wrong and those who think so are right-wing bigots who are seeking to impose their narrow-minded views on all the rest of us!

Sally and Bob are *having an argument* in this exchange. That is, they are each expressing conflicting views in a heated manner. However, that is not the sense of "argument" with which logic is concerned. Logic concerns a different sense of the word "argument." An argument, in this sense, is *a reason for thinking that a statement, claim or idea is true*. For example:

> Sally: Abortion is morally wrong because it is wrong to take the life of an innocent human being, and a fetus is an innocent human being.

In this example Sally has given an argument against the moral permissibility of abortion. That is, she has given us a reason for thinking that abortion is morally wrong. The **conclusion** of the argument is the first four words, "abortion is morally wrong." But whereas in the first example Sally was simply asserting *that* abortion

is wrong (and then trying to put down those who support it), in this example she is offering a reason for *why* abortion is wrong.

We can (and should) be more precise about our definition of an argument. But before we can do that, we need to introduce some further terminology that we will use in our definition. As I've already noted, the conclusion of Sally's argument is that abortion is morally wrong. But the reason for thinking the conclusion is true is what we call the **premise**. So we have two parts of an argument: the premise and the conclusion. Typically, a conclusion will be supported by two or more premises. Both premises and conclusions are statements. A **statement** is a type of sentence that can be true or false and corresponds to the grammatical category of a "declarative sentence." For example, the sentence,

>The Nile is a river in northeastern Africa

is a statement. Why? Because it makes sense to inquire whether it is true or false. (In this case, it happens to be true.) But a sentence is still a statement even if it is false. For example, the sentence,

>The Yangtze is a river in Japan

is still a statement; it is just a false statement (the Yangtze River is in China). In contrast, none of the following sentences are statements:

>Please help yourself to more casserole
>Don't tell your mother about the surprise
>Do you like Vietnamese pho?

The reason that none of these sentences are statements is that it doesn't make sense to ask whether those sentences are true or false (rather, they are requests or commands, and questions, respectively).

So, to reiterate: all arguments are composed of premises and conclusions, which are both types of statements. The premises of the argument provide a reason for thinking that the conclusion is true. And arguments typically involve more than one premise.

10. Validity and Soundness

1.7 Soundness

A good argument is not only valid, but also sound. Soundness is defined in terms of validity, so since we have already defined validity, we can now rely on it to define soundness. A **sound argument** is a valid argument that has all true premises. That means that the conclusion of a sound argument will always be true. Why? Because if an argument is valid, the premises transmit truth to the conclusion on the assumption of the truth of the premises. But if the premises are actually true, as they are in a sound argument, then since all sound arguments are valid, we know that the conclusion of a sound argument is true. Compare the last two Obama examples from the previous section. While the first argument was sound, the second argument was not sound, although it was valid. The relationship between soundness and validity is easy to specify: all sound arguments are valid arguments, but not all valid arguments are sound arguments.

Although soundness is what any argument should aim for, we will not be talking much about soundness in this book. The reason for this is that the only difference between a valid argument and a sound argument is that a sound argument has all true premises. But how do we determine whether the premises of an argument are actually true? Well, there are lots of ways to do that, including using Google to look up an answer, studying the relevant subjects in school, consulting experts on the relevant topics, and so on. But none of these activities have anything to do with logic, per se. The relevant disciplines to consult if you want to know whether a particular statement is true is almost never logic! For example, logic has nothing to say regarding whether or not protozoa are animals or whether there are predators that aren't in the animal kingdom.

In order to learn whether those statements are true, we'd have to consult biology, not logic. Since this is a logic textbook, however, it is best to leave the question of what is empirically true or false to the relevant disciplines that study those topics. And that is why the issue of soundness, while crucial for any good argument, is outside the purview of logic.

1.8 Deductive vs. Inductive arguments

The concepts of validity and soundness that we have introduced apply only to the class of what are called "deductive arguments". A **deductive argument** is an argument whose conclusion is supposed to follow from its premises with absolute certainty, thus leaving no possibility that the conclusion doesn't follow from the premises. For a deductive argument to fail to do this is for it to fail as a deductive argument. In contrast, an **inductive argument** is an argument whose conclusion is supposed to follow from its premises with a high level of probability, which means that although it is possible that the conclusion doesn't follow from its premises, it is unlikely that this is the case. Here is an example of an inductive argument:

> Tweets is a healthy, normally functioning bird and since most healthy, normally functioning birds fly, Tweets probably flies.

Notice that the conclusion, Tweets probably flies, contains the word "probably." This is a clear indicator that the argument is supposed to be inductive, not deductive. Here is the argument in standard form:

1. Tweets is a healthy, normally functioning bird
2. Most healthy, normally functioning birds fly
3. Therefore, Tweets probably flies

Given the information provided by the premises, the conclusion

does seem to be well supported. That is, the premises do give us a strong reason for accepting the conclusion. This is true even though we can imagine a scenario in which the premises are true and yet the conclusion is false. For example, suppose that we added the following premise:

Tweets is 6 ft tall and can run 30 mph.

Were we to add that premise, the conclusion would no longer be supported by the premises, since any bird that is 6 ft tall and can run 30 mph, is not a kind of bird that can fly. That information leads us to believe that Tweets is an ostrich or emu, which are not kinds of birds that can fly. As this example shows, inductive arguments are **defeasible arguments** since by adding further information or premises to the argument, we can overturn (defeat) the verdict that the conclusion is well-supported by the premises. Inductive arguments whose premises give us a strong, even if defeasible, reason for accepting the conclusion are called, unsurprisingly, **strong inductive arguments**. In contrast, an inductive argument that does not provide a strong reason for accepting the conclusion are called **weak inductive arguments**.

Whereas strong inductive arguments are defeasible, valid deductive arguments aren't. Suppose that instead of saying that *most* birds fly, premise 2 said that *all* birds fly.

1. Tweets is a healthy, normally function bird.
2. All healthy, normally functioning birds can fly.
3. Therefore, Tweets can fly.

This is a valid argument and since it is a valid argument, there are no further premises that we could add that could overturn the argument's validity. (True, premise 2 is false, but as we've seen that is irrelevant to determining whether an argument is valid.) Even if we were to add the premise that Tweets is 6 ft tall and can run 30 mph, it doesn't overturn the validity of the argument. As soon as we use the **universal generalization**, "*all* healthy, normally function birds can fly," then when we assume that premise is true and add

that Tweets is a healthy, normally functioning bird, it *has* to follow from those premises that Tweets can fly. This is true even if we add that Tweets is 6 ft tall because then what we have to imagine (in applying our informal test of validity) is a world in which all birds, including those that are 6 ft tall and can run 30 mph, can fly.

Although inductive arguments are an important class of argument that are commonly used every day in many contexts, logic texts tend not to spend as much time with them since we have no agreed upon standard of evaluating them. In contrast, there is an agreed upon standard of evaluation of deductive arguments. We have already seen what that is; it is the concept of validity. In chapter 2 we will learn some precise, formal methods of evaluating deductive arguments. There are no such agreed upon formal methods of evaluation for inductive arguments. This is an area of ongoing research in philosophy. In chapter 3 we will revisit inductive arguments and consider some ways to evaluate inductive arguments.

1.9 Arguments with missing premises

Quite often, an argument will not explicitly state a premise that we can see is needed in order for the argument to be valid. In such a case, we can supply the premise(s) needed in order so make the argument valid. Making missing premises explicit is a central part of reconstructing arguments in standard form. We have already dealt in part with this in the section on paraphrasing, but now that we have introduced the concept of validity, we have a useful tool for knowing when to supply missing premises in our reconstruction of an argument. In some cases, the missing premise will be fairly obvious, as in the following:

>   Gary is a convicted sex-offender, so Gary is not allowed to work with children.

The premise and conclusion of this argument are straightforward:

1. Gary is a convicted sex-offender
2. Therefore, Gary is not allowed to work with children (from 1)

However, as stated, the argument is invalid. (Before reading on, see if you can provide a counterexample for this argument. That is, come up with an imaginary scenario in which the premise is true and yet the conclusion is false.) Here is just one counterexample (there could be many): Gary is a convicted sex-offender but the country in which he lives does not restrict convicted sex-offenders from working with children. I don't know whether there are any such countries, although I suspect there are (and it doesn't matter for the purpose of validity whether there are or aren't). In any case, it seems clear that this argument is relying upon a premise that isn't explicitly stated. We can and should state that premise explicitly in our reconstruction of the standard form argument. But what is the argument's missing premise? The obvious one is that no sex-offenders are allowed to work with children, but we could also use a more carefully statement like this one:

> *Where Gary lives*, no convicted sex-offenders are allowed to work with children.

It should be obvious why this is a more "careful" statement. It is more careful because it is not so universal in scope, which means that it is easier for the statement to be made true. By relativizing the statement that sex-offenders are not allowed to work with children to the *place* where Gary lives, we leave open the possibility that other places in the world don't have this same restriction. So even if there are other places in the world where convicted sex-offenders are allowed to work with children, our statements could still be true since in *this* place (the place where Gary lives) they aren't. (For more on strong and weak statements, see section 1.10). So here is the argument in standard form:

1. Gary is a convicted sex-offender.
2. Where Gary lives, no convicted sex-offenders are allowed to work with children.

3. Therefore, Gary is not allowed to work with children. (from 1-2)

This argument is now valid: there is no way for the conclusion to be false, assuming the truth of the premises. This was a fairly simple example where the missing premise needed to make the argument valid was relatively easy to see. As we can see from this example, a **missing premise** is a premise that the argument needs in order to be as strong as possible. Typically, this means supplying the statement(s) that are needed to make the argument valid. But in addition to making the argument valid, we want to make the argument plausible. This is called "the principle of charity." The **principle of charity** states that when reconstructing an argument, you should try to make that argument (whether inductive or deductive) as strong as possible. When it comes to supplying missing premises, this means supplying the most plausible premises needed in order to make the argument either valid (for deductive arguments) or inductively strong (for inductive arguments).

Although in the last example figuring out the missing premise was relatively easy to do, it is not always so easy. Here is an argument whose missing premises are not as easy to determine:

> Since children who are raised by gay couples often have psychological and emotional problems, the state should discourage gay couples from raising children.

The conclusion of this argument, that the state should not allow gay marriage, is apparently supported by a single premise, which should be recognizable from the occurrence of the premise indicator, "since." Thus, our initial reconstruction of the standard form argument looks like this:

1. Children who are raised by gay couples often have psychological and emotional problems.
2. Therefore, the state should discourage gay couples from raising children.

However, as it stands, this argument is invalid because it depends

on certain missing premises. The conclusion of this argument is a normative statement—a statement about whether something *ought to be true*, relative to some standard of evaluation. Normative statements can be contrasted with descriptive statements, which are simply factual claims about what *is true*. For example, "Russia *does not* allow gay couples to raise children" is a descriptive statement. That is, it is simply a claim about what is *in fact the case* in Russia today. In contrast, "Russia *should not* allow gay couples to raise children" is a normative statement since it is not a claim about what is true, but what ought to be true, relative to some standard of evaluation (for example, a moral or legal standard). An important idea within philosophy, which is often traced back to the Scottish philosopher David Hume (1711-1776), is that statements about what ought to be the case (i.e., normative statements) can never be derived from statements about what is the case (i.e., descriptive statements). This is known within philosophy as the is-ought gap. The problem with the above argument is that it attempts to infer a normative statement from a purely descriptive statement, violating the is-ought gap. We can see the problem by constructing a counterexample. Suppose that in society x it is true that children raised by gay couples have psychological problems. However, suppose that in that society people do not accept that the state should do what it can to decrease harm to children. In this case, the conclusion, that the state should discourage gay couples from raising children, does not follow. Thus, we can see that the argument depends on a missing or assumed premise that is not explicitly stated. That missing premise must be a normative statement, in order that we can infer the conclusion, which is also a normative statement. There is an important general lesson here: *Many times an argument with a normative conclusion will depend on a normative premise which is not explicitly stated.* The missing normative premise of this particular argument seems to be something like this:

The state should always do what it can to decrease harm to children.

Notice that this is a normative statement, which is indicated by the use of the word "should." There are many other words that can be used to capture normative statements such as: good, bad, and ought. Thus, we can reconstruct the argument, filling in the missing normative premise like this:

1. Children who are raised by gay couples often have psychological and emotional problems.
2. The state should always do what it can to decrease harm to children.
3. Therefore, the state should discourage gay couples from raising children. (from 1-2)

However, although the argument is now in better shape, it is *still* invalid because it is still possible for the premises to be true and yet the conclusion false. In order to show this, we just have to imagine a scenario in which both the premises are true and yet the conclusion is false. Here is one counterexample to the argument (there are many). Suppose that while it is true that children of gay couples often have psychological and emotional problems, the rate of psychological problems in children raised by gay couples is actually lower than in children raised by heterosexual couples. In this case, even if it were true that the state should always do what it can to decrease harm to children, it does not follow that the state should discourage gay couples from raising children. In fact, in the scenario I've described, just the opposite would seem to follow: the state should discourage heterosexual couples from raising children.

But even if we suppose that the rate of psychological problems in children of gay couples is higher than in children of heterosexual couples, the conclusion still doesn't seem to follow. For example, it could be that the reason that children of gay couples have higher rates of psychological problems is that in a society that is not yet accepting of gay couples, children of gay couples will face more teasing, bullying and general lack of acceptance than children of heterosexual couples. If this were true, then the harm to these

children isn't so much due to the fact that their parents are gay as it is to the fact that their community does not accept them. In that case, the state should not necessarily discourage gay couples from raising children. Here is an analogy: At one point in our country's history (if not still today) it is plausible that the children of black Americans suffered more psychologically and emotionally than the children of white Americans. But for the government to discourage black Americans from raising children would have been unjust, since it is likely that if there was a higher incidence of psychological and emotional problems in black Americans, then it was due to unjust and unequal conditions, not to the black parents, per se. So, to return to our example, the state should only discourage gay couples from raising children if they know that the higher incidence of psychological problems in children of gay couples *isn't* the result of any kind of injustice, but is due to the simple fact that the parents are gay.

Thus, one way of making the argument (at least closer to) valid would be to add the following two missing premises:

> A. The rate of psychological problems in children of gay couples is higher than in children of heterosexual couples.
>
> B. The higher incidence of psychological problems in children of gay couples is not due to any kind of injustice in society, but to the fact that the parents are gay.

So the reconstructed standard form argument would look like this:

1. Children who are raised by gay couples often have psychological and emotional problems.
2. The rate of psychological problems in children of gay couples is higher than in children of heterosexual couples.
3. The higher incidence of psychological problems in children of gay couples is not due to any kind of injustice in society, but to the fact that the parents are gay.
4. The state should always do what it can to decrease harm to children.

5. Therefore, the state should discourage gay couples from raising children. (from 1-4)

In this argument, premises 2-4 are the missing or assumed premises. Their addition makes the argument much stronger, but making them explicit enables us to clearly see what assumptions the argument relies on in order for the argument to be valid. This is useful since we can now clearly see which premises of the argument we may challenge as false. Arguably, premise 4 is false, since the state shouldn't always do what it can to decrease harm to children. Rather, it should only do so as long as such an action didn't violate other rights that the state has to protect or create larger harms elsewhere.

The important lesson from this example is that supplying the missing premises of an argument is not always a simple matter. In the example above, I have used the principle of charity to supply missing premises. Mastering this skill is truly an art (rather than a science) since there is never just one correct way of doing it (cf. section 1.5) and because it requires a lot of skilled practice.

> Exercise 6: Supply the missing premise or premises needed in order to make the following arguments valid. Try to make the premises as plausible as possible while making the argument valid (which is to apply the principle of charity).

1. Ed rides horses. Therefore, Ed is a cowboy.
2. Tom was driving over the speed limit. Therefore, Tom was doing something wrong.
3. If it is raining then the ground is wet. Therefore, the ground must be wet.
4. All elves drink Guinness, which is why Olaf drinks Guinness.
5. Mark didn't invite me to homecoming. Instead, he invited his friend Alexia. So he must like Alexia more than me.
6. The watch must be broken because every time I have looked at it, the hands have been in the same place.

7. Olaf drank too much Guinness and fell out of his second story apartment window. Therefore, drinking too much Guinness caused Olaf to injure himself.
8. Mark jumped into the air. Therefore, Mark landed back on the ground.
9. In 2009 in the United States, the net worth of the median white household was $113,149 a year, whereas the net worth of the median black household was $5,677. Therefore, as of 2009, the United States was still a racist nation.
10. The temperature of the water is 212 degrees Fahrenheit. Therefore, the water is boiling.
11. Capital punishment sometimes takes innocent lives, such as the lives of individuals who were later found to be not guilty. Therefore, we should not allow capital punishment.
12. Allowing immigrants to migrate to the U.S. will take working class jobs away from working class folks. Therefore, we should not allow immigrants to migrate to the U.S.
13. Prostitution is a fair economic exchange between two consenting adults. Therefore, prostitution should be allowed.
14. Colleges are more interested in making money off of their football athletes than in educating them. Therefore, college football ought to be banned.
15. Edward received an F in college Algebra. Therefore, Edward should have studied more.

11. Attribution

The material in this chapter is adapted from previously published content:

Van Cleave, M.J. (2017). *Introduction to Logic and Critical Thinking.* Lansing Community College. Section 1.1, 1.17-23. https://drive.google.com/file/d/0B-rkbctcWjo3OFhGODNSYTZPMVU/ view?resourcekey=0-iGfOufS0tNl2MRQa64ITng CC BY.

TESTING ARGUMENTS

12. Formal vs. Informal Fallacies

Formal vs. Informal Fallacies

A **fallacy** is simply a mistake in reasoning. Some fallacies are formal and some are informal. In chapter 2, we saw that we could define validity formally and thus could determine whether an argument was valid or invalid without even having to know or understand what the argument was about. We saw that we could define certain valid rules of inference, such as modus ponens and modus tollens. These inference patterns are valid in virtue of their form, not their content. That is, any argument that has the same form as modus ponens or modus tollens will automatically be valid. A formal fallacy is simply an argument whose form is invalid. Thus, any argument that has that form will automatically be invalid, regardless of the meaning of the sentences. Two formal fallacies that are similar to, but should never be confused with, modus ponens and modus tollens are **denying the antecedent** and **affirming the consequent**. Here are the forms of those invalid inferences:

Denying the antecedent

p ⊃ q

~p

∴ ~q

Affirming the consequent

p ⊃ q

q

∴ p

Any argument that has either of these forms is an invalid argument. For example:

1. If Kant was a deontologist, then he was a non-consequentialist.
2. Kant was not a deontologist.
3. Therefore, Kant was a not a non-consequentialist.

The form of this argument is:

1. D ⊃ C
2. ~D
3. ∴ ~C

As you can see, this argument has the form of the fallacy, denying the antecedent. Thus, we know that this argument is invalid even if we don't know what "Kant" or "deontologist" or "non-consequentialist" means. ("Kant" was a famous German philosopher from the early 1800s, whereas "deontology" and "non-consequentialist" are terms that come from ethical theory.) It is mark of a formal fallacy that we can identify it even if we don't really understand the meanings of the sentences in the argument. Recall our Jabberwocky argument from chapter 2. Here's an argument which uses silly, made-up words from Lewis Carrol's "Jabberwocky." See if you can determine whether the argument's form is valid or invalid:

1. If toves are brillig then toves are slithy.
2. Toves are slithy
3. Therefore, toves are brillig.

You should be able to see that this argument has the form of affirming the consequent:

1. B ⊃ S
2. S
3. ∴ B

As such, we know that the argument is invalid, even though we haven't got a clue what "toves" are or what "slithy" or "brillig" means.

The point is that we can identify formal fallacies without having to know what they mean.

In contrast, **informal fallacies** are those which cannot be identified without understanding the concepts involved in the argument. A paradigm example of an informal fallacy is the fallacy of composition. We will consider this fallacy in the next sub-section. In the remaining subsections, we will consider a number of other informal logical fallacies.

4.1.1 Composition fallacy

Consider the following argument:

> Each member on the gymnastics team weighs less than 110 lbs. Therefore, the whole gymnastics team weighs less than 110 lbs.

This arguments commits the composition fallacy. In the composition fallacy one argues that since each part of the whole has a certain feature, it follows that the whole has that same feature. However, you cannot generally identify any argument that moves from statements about parts to statements about wholes as committing the composition fallacy because whether or not there is a fallacy depends on what feature we are attributing to the parts and wholes. Here is an example of an argument that moves from claims about the parts possessing a feature to a claim about the whole possessing that same feature, but doesn't commit the composition fallacy:

> Every part of the car is made of plastic. Therefore, the whole car is made of plastic.

This conclusion *does* follow from the premises; there is no fallacy here. The difference between this argument and the preceding argument (about the gymnastics team) isn't their form. In fact both arguments have the same form:

> Every part of X has the feature f. Therefore, the whole X
> has the feature f.

And yet one of the arguments is clearly fallacious, while the other isn't. The difference between the two arguments is not their form, but their content. That is, the difference is what feature is being attributed to the parts and wholes. Some features (like *weighing a certain amount*) are such that if they belong to each part, then it does not follow that they belong to the whole. Other features (such as *being made of plastic*) are such that if they belong to each part, it follows that they belong to the whole.

Here is another example:

> Every member of the team has been to Paris. Therefore
> the team has been to Paris.

The conclusion of this argument does not follow. Just because each member of the team has been to Paris, it doesn't follow that the whole team has been to Paris, since it may not have been the case that each individual was there at the same time and was there in their capacity as a member of the team. Thus, even though it is plausible to say that the team is composed of every member of the team, it doesn't follow that since every member of the team has been to Paris, the whole team has been to Paris. Contrast that example with this one:

> Every member of the team was on the plane. Therefore,
> the whole team was on the plane.

This argument, in contrast to the last one, contains no fallacy. It is true that if every member is on the plane then the whole team is on the plane. And yet these two arguments have almost exactly the same form. The only difference is that the first argument is talking about the property, *having been to Paris*, whereas the second argument is talking about the property, *being on the plane*. The only reason we are able to identify the first argument as committing the composition fallacy and the second argument as not committing a fallacy is that we understand the relationship between the concepts involved. In the first case, we understand that it is possible that every member could have been to Paris without the team ever

having been; in the second case we understand that as long as every member of the team is on the plane, it has to be true that the whole team is on the plane. The take home point here is that in order to identify whether an argument has committed the composition fallacy, one must understand the concepts involved in the argument. This is the mark of an informal fallacy: we have to rely on our understanding of the meanings of the words or concepts involved, rather than simply being able to identify the fallacy from its form.

13. Constructing Arguments

Consider the following argument:

1. All humans are mortal
2. All mortal things die
3. Therefore, all humans die

Categorical logic is the logic that deals with the logical relationship between categorical statements. A **categorical statement** is simply a statement about a category or type of thing. For example, the first premise of the above argument is a statement about the categories of humans and things that are mortal. The second premise is a statement about the categories of things that are mortal and things that die. Finally, the conclusion is a statement about humans and things that die. Although you may think that this argument as a similar form as a hypothetical syllogism, it is distinct from a hypothetical syllogism because the premises are not composed of two different atomic propositions. Rather, each premise contains only one atomic proposition.

In categorical logic, the logical terms (analogous to the truth functional operators of propositional logic) are the terms "all" and "some." In contrast with propositional logic, in categorical logic we will use capital letters to stand for categories of things in the world, rather than for atomic propositions. Thus, we can represent the statement:

All humans are mortal

as

All H are M

where "H" stands for the category of "humans" and "M" stands for the category, "things that are mortal." Notice that the categories are nouns or noun phrases. Thus, instead of saying that the category is "mortal" I said the category is "things that are mortal." It is important to recognize the difference between how the capital letters are

being used in categorical logic and how they were used in propositional logic. In categorical logic, the capital letters stand for noun phrases that denote categories of things in the world—for example, "cars" or "things that are man-made" or "mammals" or "things that are red."

   In categorical logic, we will use what are called **Venn diagrams** to represent the logical relationships between the different kinds of categorical statements. A Venn diagram is simply a way of graphically representing the logical relationship between two different categorical statements. Below is a Venn diagram that represents the statement, "all humans are mortal."
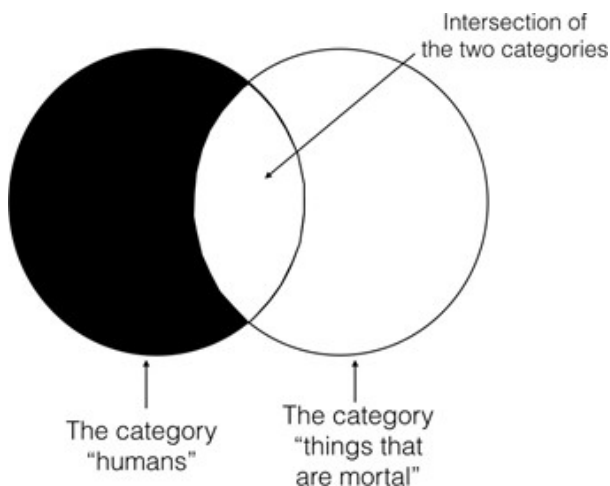
Here is how to understand this Venn. There are two circles that represent the two categories, "humans" and "things that are mortal." These two categories are overlapping so that the intersection of those two categories (i.e. the place where the two circles overlap) represents things that are both human and mortal. Any shaded portions of the Venn diagram (by "shaded" I will mean "blacked out") represent that there is nothing in that area of the

category. So the above Venn says that there is nothing in the category "humans" that is not also in the category "things that are mortal." The above Venn also allows that there are things that are in the category "things that are mortal" but that aren't in the category "humans" (which is as it should be since, of course, dogs are mortal and yet not human). So the reason the category "things that are mortal" is left unshaded is that in saying "all humans are mortal" I leave open the possibility that there are things that are not human and yet mortal.

As noted above, the statement, "all humans are mortal," has a particular form:

All H are M.

This is one of the **four categorical forms**. The way we will represent these categorical forms generally are with an "S" (which stands for "subject term") and a "P" (which stands for "predicate term"). Thus, the categorical statement, "all humans are mortal," has the following categorical form:

All S are P

The way we interpret statements of this form are as follows: everything in the category S is also in the category P. This statement form is what we call a "**universal affirmative**," since it is a universal statement that does not contain a negation. There are three other categorical statement forms that you will have to become familiar with in order to do categorical logic. Here they are (with the name of the type of statement in parentheses to the right:

No S are P   (**universal negative**)
Some S are P   (**particular affirmative**)
Some S are not P   (**particular negative**)

Here are three examples of statements that have these three forms (respectively):

No reptiles give live birth

Some birds are taller than President Obama
Some birds don't fly

Notice that although these three statements don't have exactly the same form as the statement forms above, they can be translated into those same forms. All we have to do is figure out the noun phrase that describes each category that the statement is referring to. Let's start with "no reptiles give live birth." This categorical statement refers to two different categories: the category of "reptiles" and the category of "things that give live birth." Notice, again, that I added "things that…" to the predicate of the sentence ("give live birth") because "give live birth" is not a description of a category. Rather, the way of describing the category is with the noun phrase, "things that give live birth." Using these two category descriptions, we can translate this sentence to have the same form as its categorical form. All we have to do is substitute in the name of the subject category (i.e., the "S" term) and the description of the predicate category (i.e., the "P" term). Doing that will yield the following sentence:

No *reptiles* are *things that give live birth*

Although this sentence sounds strange in English, it has the same form as the categorical form, no S are P, and this translation allows us to clearly see that it does and thus to see what the two categories are. Here is what the Venn diagram for this statement looks like:

Reptiles   Things that
                give live birth
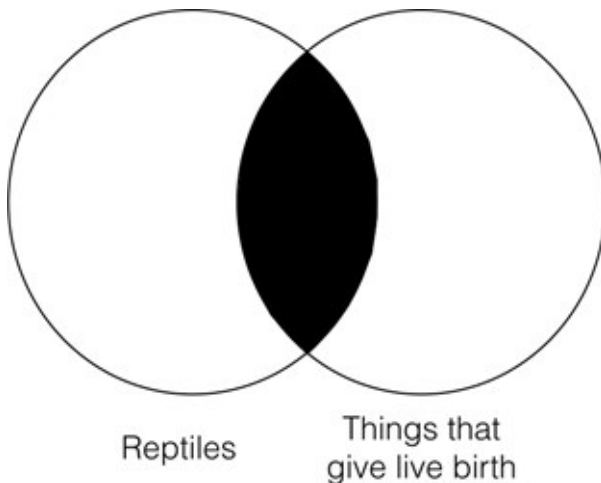
This Venn diagram represents that there is nothing in the intersection of the two categories, "reptiles" and "things that give live birth." If you think about it, this is exactly what our original statement was saying: there isn't anything that is both a reptile and gives live birth.

Let's look at the next statement, "some birds are taller than President Obama." This is a statement not about all birds, but about some birds. What are the two categories? One category is clearly "birds." The other category is "things that are taller than President Obama." That may sound like a strange category, but it is perfectly legitimate category. It includes things like adult ostriches, large grizzly bears standing on their hind legs, giraffes, the Flatiron Building, a school bus, etc. Here is how we'd translate this sentence using our two categories:

> Some birds are things that are taller than President Obama.

Again, although this sentence sounds strange in English, it has the same form as the categorical form, some S are P, and it allows us to clearly see what the two categories are. Below is the Venn diagram for this statement:

Birds   Things that are taller
               than President Obama
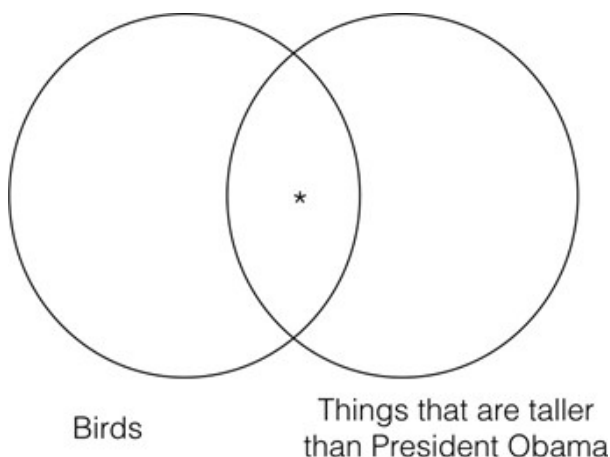
By convention, an asterisk on the Venn diagram means that there is at least one thing in that category. By putting the asterisk in the intersection of the two categories, we are saying that there is at least one thing that is a bird and is taller than President Obama, which is exactly what our original sentence was saying.

Finally, let's consider the statement, "some birds don't fly." How would we translate this sentence to have the "some S are not P" form? The first step is to

get the descriptions of the two categories using either nouns or noun phrases. The "S" term is easy; it is just "birds" again. But we have to be a bit more careful with the "P" term, since its predicate contains a negation. We do not want any of our categories to contain a negation. Rather, the negation is contained in the form (i.e., the "not"). The category cannot be simply "fly" or even "flies" since neither of these are a category of *thing*. We have to use our trick of turning the predicate into a noun phrase, i.e., "things that fly." Given these two category descriptions, we can then translate the sentence to have the categorical form, some S are not P:

Some *birds* are not *things that fly*

Again, although the English sounds clunky here, it has the same form as the categorical form, some S are not P, and it allows us to clearly see what the two categories are. Below is the Venn diagram for this statement:
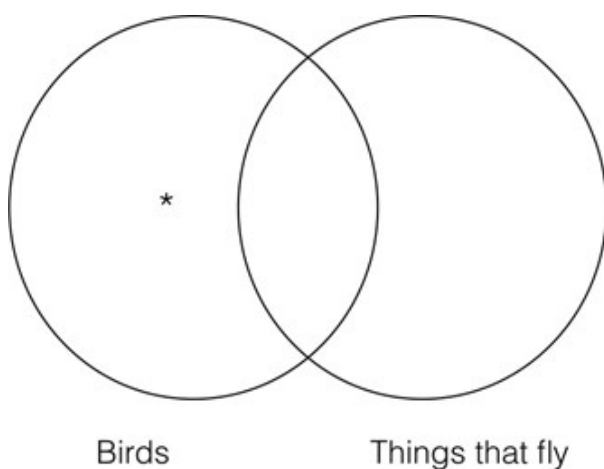
By convention, an asterisk on the Venn diagram means that there is at least one thing in that category. By putting the asterisk inside the "birds" category, but outside the "things that fly" category, we are representing that at least one thing that is a bird isn't a thing that flies. This is exactly what our original sentence was saying.

Translating categorical statements into their categorical form can by tricky. In fact, it is probably one of the trickier things you'll do in formal logic. There is no simple way of doing it other than asking yourself whether your translation accurately captures the meaning of the original English sentence. Here is an example of a tricky categorical statement:

Nobody loves me but my mother.

This is a categorical statement, but which of the four categorical forms does it have? The first step is to ask what two categories

are being referred to in this sentence. Here are the two categories: "things that love me" and "things that are my mother." Notice that the category couldn't just be "my mother" since that isn't a category; it's a particular thing. Again, this sounds strange, but it is important to remember that we are describing *categories of things*. The next question is: what is this sentence saying is the relationship between these two categories? Hint: it has to be one of the four categorical forms (since any categorical statement can be translated into one of these four forms). The sentence is saying that the only things that love me are things that are my mother. The categorical form of the statement is the "all S are P" form. Thus, the sentence, translated into the correct categorical form would be:

All *things that love me* are *things that are my mother*.

We will end this section with one last example. Consider the following categorical statement:

The baboon is a fearsome beast.

Which of the four categorical forms does this statement have? Although the article "the," which often denotes particulars, may lead one to think that this is a particular affirmative form (some S are P), it is actually a universal affirmative form (all S are P). This English sentence has the sense of "baboons are fearsome beasts" rather than of "that (particular) baboon is a fearsome beast." English is strange, which is what makes translation one of the trickiest parts of logic. So, the two categories are: "baboons" and "fearsome beasts." Notice that since "fearsome beasts" is already a noun phrase, we don't have to add "things that are..." to it. Using the two category descriptions, the translation into the "all S are P" categorical form is thus:

All *baboons* are *fearsome beasts*.

In this section we have learned what categorical statement are, how to translate categorical statements into one of the four categorical forms, and how to construct Venn diagrams for each of the four categorical forms. The following exercises will give you some practice with the translation part; in subsequent sections

we will learn how to use Venn diagrams as a formal method of evaluating a certain class of arguments.

> Exercise: Translate each of the following sentences into one of the four categorical forms (universal affirmative, universal negative, particular affirmative, particular negative). Make sure that the descriptions of the two categories are nouns or noun phrases (rather than adjectives or verbs).

1. Real men wear pink.
2. Dinosaurs are not birds.
3. Birds evolved from dinosaurs.
4. Some mammals are not predators.
5. Some predators are not mammals.
6. Not all who wander are lost.
7. All presidents are not women.
8. Boxers aren't rich.
9. If someone is sleeping then they aren't conscious.
10. If someone is conscious then they aren't sleeping.
11. All's well that ends well.
12. My friends are the only ones that care.
13. Someone loves you.
14. Jesus loves everyone.
15. Jesus loves the little children.
16. Some people don't love Jesus.
17. Only pedestrians may use the Appalachian Trail.
18. Only citizens can be president.
19. Anyone who is a Hindu believes in God.
20. Anything that is cheap is no good.
21. Some expensive things are no good.
22. Not all mammals have legs.
23. There are couples without children.
24. There are no people who hate chocolate.
25. There are people who hate cats.
26. Nothing that is sharp is safe.

27.   No poodle could run faster than a cheetah.
28.   No professional runner is slow.
29.   Baboons aren't friendly.
30.   Pigs will eat anything.

The Venn test of validity for immediate categorical inferences

In the last section, we introduced the four categorical forms. Those forms are below.

All S are P

S   P

No S are P

S   P

Some S are P

S P
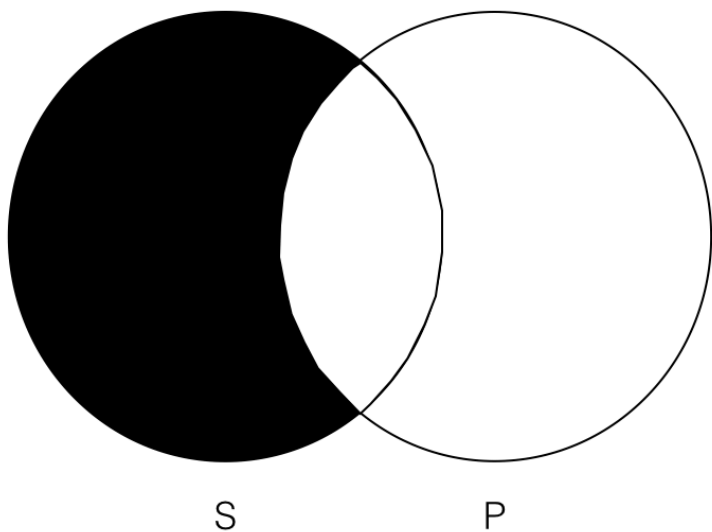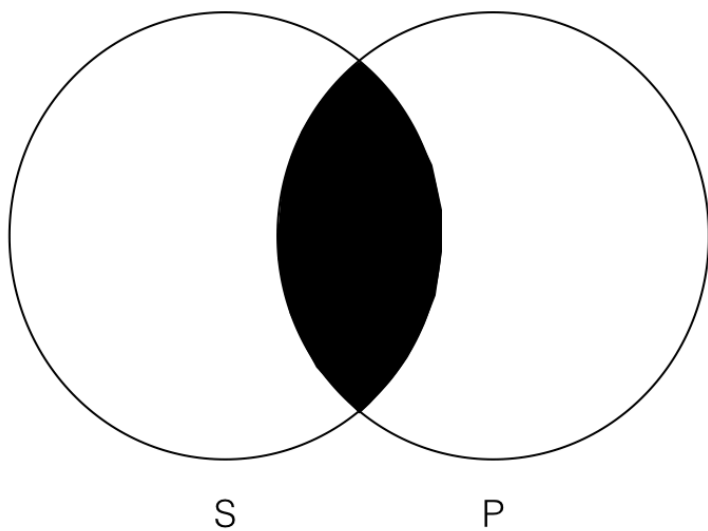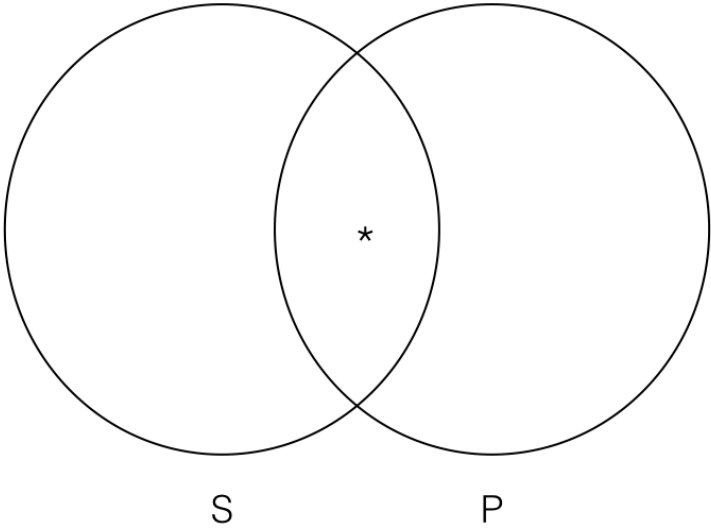
Some S are not P

S P
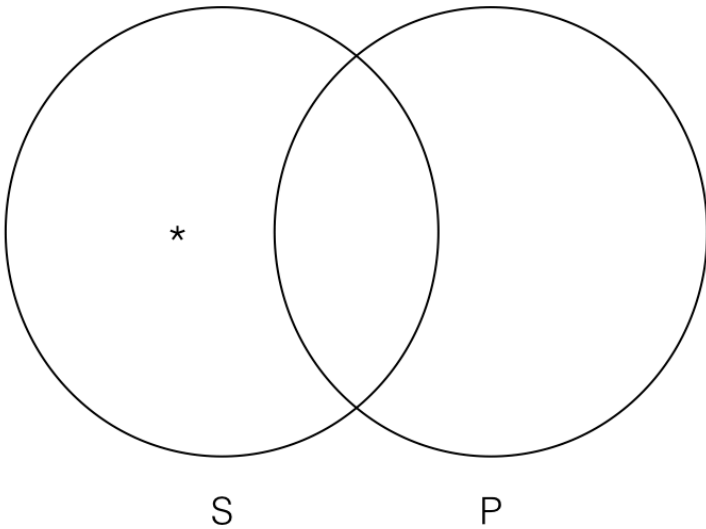
We can use Venn diagrams in order to determine whether certain kinds of arguments are valid or invalid. One such type of argument is

what we will call "immediate categorical inferences." An immediate categorical inference is simply an argument with one premise and one conclusion. For example:

1. Some mammals are amphibious.
2. Therefore, some amphibious things are mammals.

If we construct a Venn diagram for the premise and another Venn diagram for the conclusion, we will see that the Venn diagrams are identical to each other.

Mammals

Things that are amphibious

Mammals

Things that are amphibious
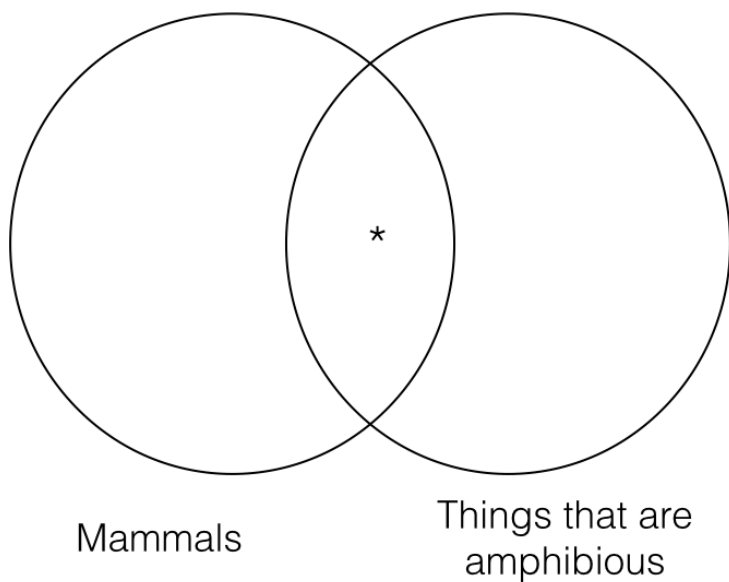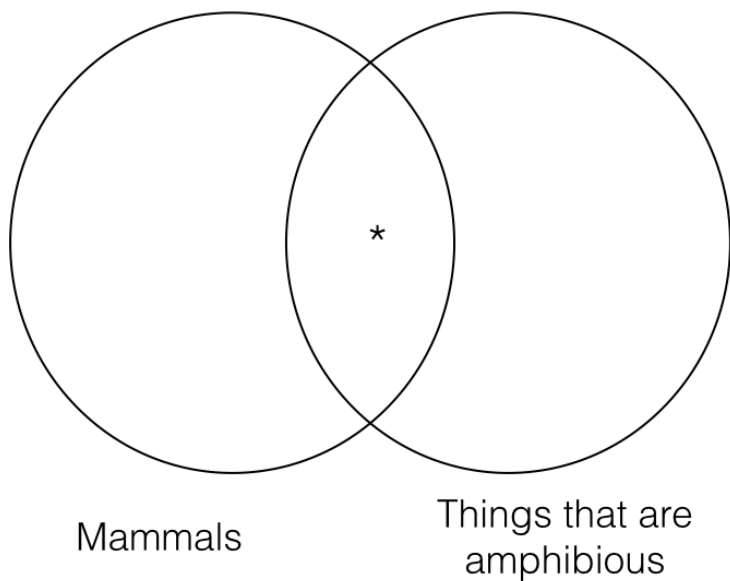
That is, the information that is represented in the Venn for the premise, is exactly the same information represented in the Venn for the conclusion. This argument passes the **Venn test of validity**

because the conclusion Venn contains no additional information that is not already contained in the premise Venn. Thus, this argument is valid. Let's now turn to an example of an invalid argument.

1. All cars are vehicles.
2. Therefore, all vehicles are cars.

Here are the Venns for the premise and the conclusion, respectively:

Cars   Vehicles

Cars   Vehicles
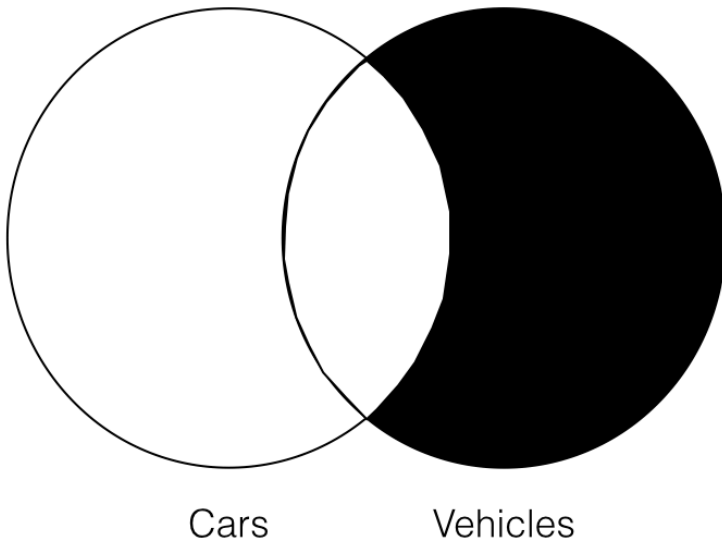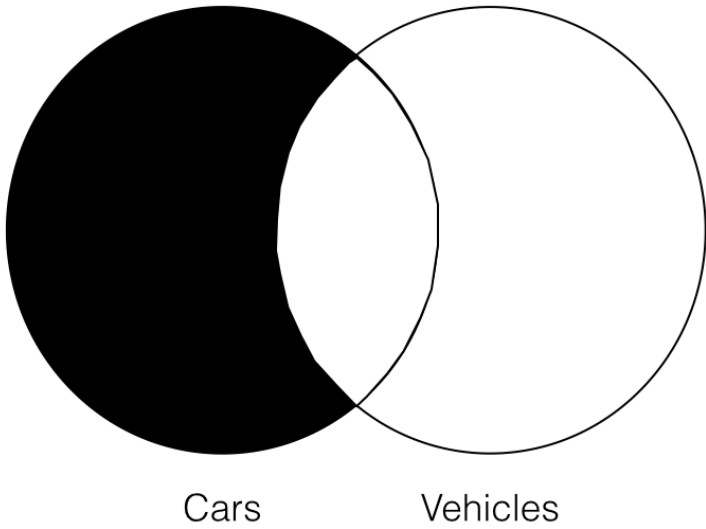
In this case, the Venns are clearly not the same. More importantly, we can see that the conclusion Venn (on the right) contains additional information that is not already contained in the premise Venn. In particular, the conclusion Venn allows that a) there could

be things in the "car" category that aren't in the "vehicle" category and b) that there cannot be anything in the "vehicle" category that isn't also in the "car" category. That is not information that is contained in the premise Venn, which says that a) there isn't anything in the category "car" that isn't also in the category "vehicle" and b) that there could be things in the category "vehicle" that aren't in the category "car." Thus, this argument does not pass the Venn test of validity since there is information contained in the conclusion Venn that is not already contained in the premise Venn. Thus, this argument is invalid.

The Venn test of validity is a formal method, because we can apply it even if we only know the form of the categorical statements, but don't know what the categories referred to in the statements represent. For example, we can simply use "S" and "P" for the categories—and we clearly don't know what these represent. For example:

1. All S are P
2. No P are S

S   P

S   P
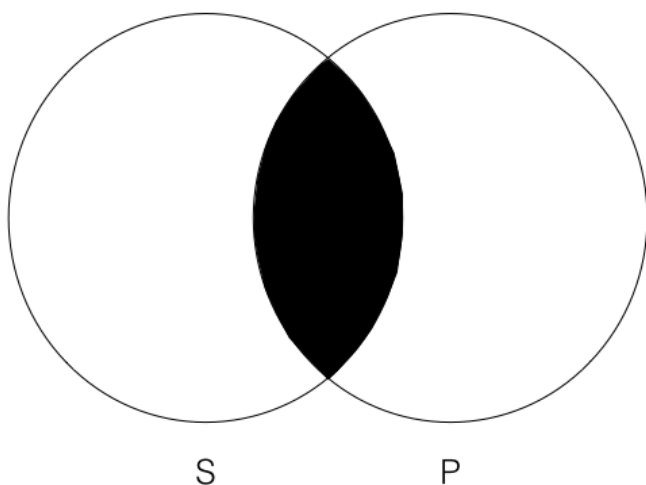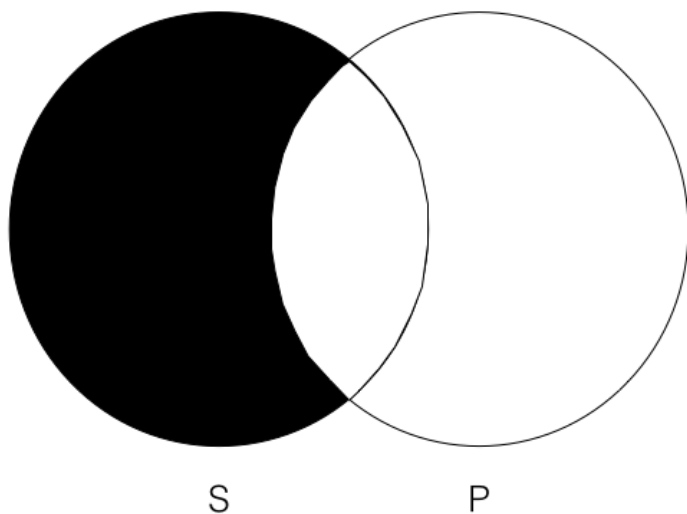
The conclusion (on the right) contains information that is not contained in the premise (on the left). In particular, the conclusion Venn explicitly rules out that there is anything that is both in the category "S" and in the category "P" while the premise Venn allows that this is the case (but does not require it). Thus, we can say that this argument fails the Venn test of validity and thus is invalid. We

know this even though we have no idea what the categories "S" and "P" are.

This is the mark of a formal method of evaluation.

> Exercise: Apply the Venn test of validity in order to determine whether the following categorical inferences are valid or invalid.

1. All S are P; therefore, all P are S
2. Some S are P; therefore, some P are S
3. Some S are P; therefore, some P are not S
4. Some S are P; therefore, all P are S
5. No S are P; therefore, no P are S
6. No P are S; therefore, some S are P
7. Some S are not P; therefore, some P are not S
8. All S are P; therefore some P are not S

Universal statements and existential commitment

Consider the following inference:

1. All S are P
2. Therefore, some S are P

Is this inference valid or invalid? As it turns out, this is an issue on which there has been much philosophical debate. On the one hand, it seems that many times when we make a universal statement, such as "all dogs are mammals," we imply that there *are* dogs—i.e., that dogs exist. Thus, if we assert that *all* dogs are mammals, that implies that *some* dogs are mammals (just as if I say that *everyone* at the party was drunk, this implies that at least *someone* at the party was drunk). In general, it may seem that "all" implies "some" (since some is encompassed by all). This reasoning would support the idea that the above inference is valid: universal statements imply

certain particular statements. Thus, statements of the form "all S are P" would imply that statements of the form "some S are P." This is what is called "**existential commitment**."

In contrast to the reasoning just laid out, modern logicians *reject* existential commitment; they do *not* take statements of the form "all S are P" to imply that there exists anything in the "S" category. Why would they think this? One way of understanding why universal statements are interpreted in this way in modern logic is by considering laws such as the following:

> All trespassers will be fined.
>
> All bodies that are not acted on by any force are at rest.
>
> All passenger cars that can travel 770 mph are supersonic.

The "S" terms in the above categorical statements are "trespassers," "bodies that are not acted on by any force," and "passenger cars that can travel 770 mph." Now ask yourself: do these statements commit us to the existence of either trespassers or bodies not acted on by any force? No, they don't. Just because we assert the rule that all trespassers will be fined, we do not necessarily commit ourselves to the claim that there are trespassers. Rather, what we are saying is *anything that is a trespasser* will be fined. But this can be true, even if there are no trespassers! Likewise, when Isaac Newton asserted that all bodies that are not acted on by any force remain at rest, he was not committing himself to the existence of "bodies not acted on by any force." Rather, he was saying that *anything that is a body not acted on by any force* will remain in motion. But this can be true, even if there are no bodies not acted on by any force! (And there aren't any such bodies, since even things that are stationary like your house or your car parked in the driveway are still acted on by forces such as gravity and friction.) Finally, in asserting that all passenger cars that can travel 770 mph are supersonic, we are not committing ourselves to the existence of any such car. Rather, we are only saying that *were there any such car*, it would be supersonic (i.e., it would travel faster than the speed of sound).

For various reasons (that we will not discuss here), modern logic

treats a universal categorical statement as a kind of conditional statement. Thus, a statement like,

All passenger cars that can travel 770 mph are supersonic

is interpreted as follows:

For any x, *if x is a passenger car that can travel 770 mph then x is supersonic.*

But since conditional statements do not assert either the antecedent or the consequent, the universal statement is not asserting the existence of passenger cars that can travel 770 mph. Rather, it is just saying that *if* there were passenger cars that could travel that fast, then those things would be supersonic.

We will follow modern logic in denying existential commitment. That is, we will not interpret universal affirmative statements of the form "All S are P" as implying particular affirmative statements of the form "some S are P." Likewise, we will not interpret universal negative statements of the form "no S are P" as implying particular negative statements of the form "some S are not P." Thus, when constructing Venn diagrams, you can always rely on the fact that if there is no particular represented in the premise Venn (i.e., there is no asterisk), then if the conclusion Venn represents a particular (i.e., there is an asterisk), the argument will be invalid. This is so since no universal statement logically implies the existence of any particular. Conversely, if the premise Venn does represent a particular statement (i.e., it contains an asterisk), then if the conclusion doesn't contain particular statement (i.e., doesn't contain an asterisk), the argument will be invalid.

Exercise: Construct Venn diagrams to determine which of the following immediate categorical inferences are valid and which are invalid. Make sure you remember that we are not interpreting universal statements to imply existential commitment.

1. All S are P; therefore, some S are P
2. No S are P; therefore, some S are not P
3. All S are P; therefore, some P are S

4. No S are P; therefore, some P are not S

Venn validity for categorical syllogisms

A **categorical syllogism** is just an argument with two premises and a conclusion, where every statement of the argument is a categorical statement. As we have seen, there are four different types (forms) of categorical statement:

> All S are P (universal affirmative)
> No S are P (universal negative)
> Some S are P (particular affirmative)
> Some S are not P (particular negative)

Thus, any categorical syllogism's premises and conclusion will be some mixture of these different types of statement. The argument I gave at the beginning of section 2.13 was a categorical syllogism. Here, again, is that argument:

1. All humans are mortal
2. All mortal things die
3. Therefore, all humans die

As we can see now that we have learned the four categorical forms, each one of the statements in this syllogism is a "universal affirmative" statement of the form, "all S are P." Let's first translate each statement of this argument to have the "all S are P" form:

1. All humans are things that are mortal.
2. All things that are mortal are things that die.
3. All humans are things that die.

In determining the validity of categorical syllogisms, we must construct a three category Venn diagram for the premises and a two category Venn diagram for the conclusion. Here is what the three category Venn looks like for the premises:

Humans

Things
that die

Things that are mortal
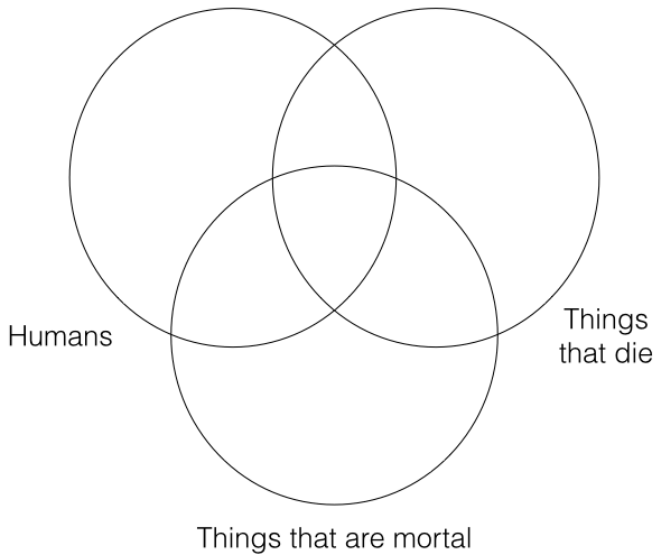
We need a three category Venn for the premises since the two premises refer to three different categories. The way you should construct the Venn is with the circle that represents the "S" category of the conclusion (i.e., the category "humans") on left, the circle that represents the "P" category of the conclusion (i.e., the category "things that die") on the right, and the remaining category ("things that are mortal") in the middle, as I have done above. Constructing your three category Venn in this way will allow you to easily determine whether the argument is valid.

The next thing we must do is represent the information from the first two premises in our three category Venn. We'll start with the first premise, which says "all humans are things that are mortal." That means that we must shade out anything that is in the "human" category, but that isn't in the "things that are mortal" category, like this:
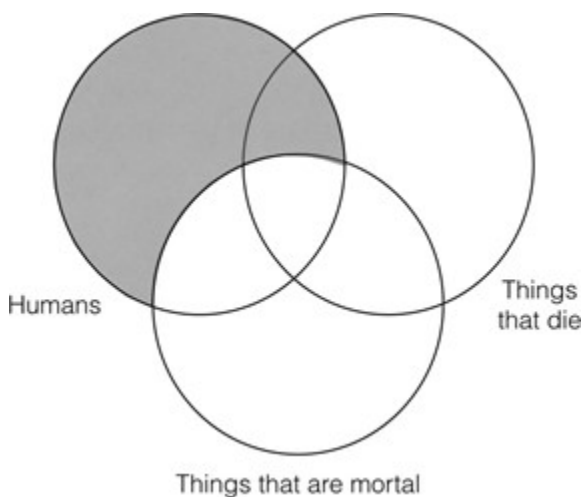
The next thing we have to do is fill in the information for the second premise, all things that are mortal are things that die. That means that there isn't anything that is in the category "things that are mortal" but that isn't in the "things that die" category. So we must shade out all of the parts of the "things that are mortal" category the lie outside the "things that die" category, like this:
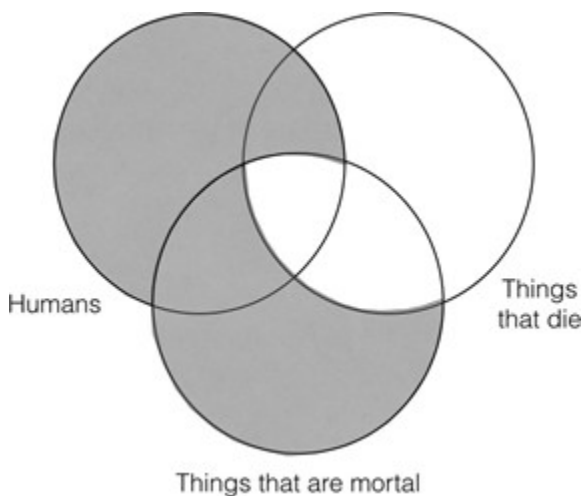
The next thing we have to do is construct a two category Venn for the conclusion and then compare the information represented by the three category Venn for the premises to the two category Venn for the conclusion.

Humans   Things that die
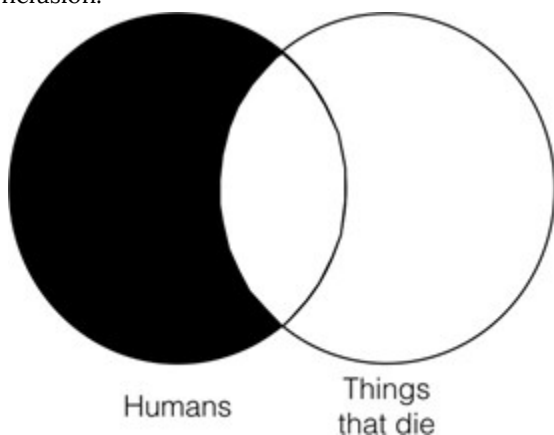
The conclusion represents the information that there is nothing in the "humans" category that isn't also in the "things that die" category. It also allows that there are things that die, but that aren't humans. The premise Venn also includes this same information, since every part of the "humans" category that is outside the "things that die" category is shaded out. Thus, this argument passes the Venn test of validity and is thus valid since there is no information represented in the conclusion Venn that is not also represented in the premise Venn. Notice that it doesn't matter that the premise Venn contains more information than the conclusion Venn. That is to be expected, since the premise Venn is representing a whole other category that the conclusion Venn isn't. This is perfectly allowable. What isn't allowable (and thus would make an argument fail the Venn test of validity) is if the conclusion Venn contained information that wasn't already contained in the premise Venn. However, since this argument does not do that, it is valid.

Let's try another one.
1. All pediatricians are doctors

2.   All pediatricians like children
3.   Therefore, all doctors like children.

The first step is to identify the three categories referred to in this categorical syllogism. They are:

Pediatricians Doctors
Things that like children

The next step is to fill out the three category Venn for the premises and the two category Venn for the conclusion.

Doctors

Things that
like children

Pediatricians
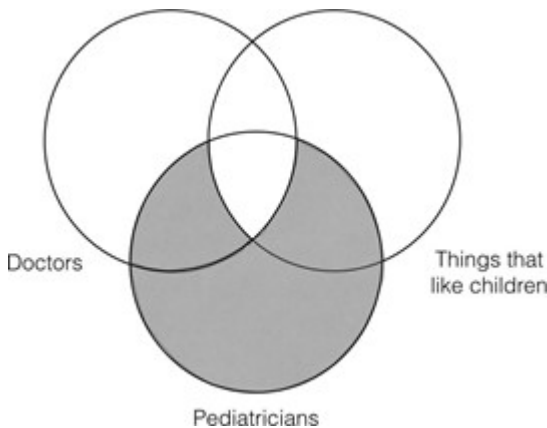
This argument does not pass the Venn test of validity because there is information contained in the conclusion Venn that is not contained in the premise Venn. In particular, the conclusion says

Doctors

Things that
like children
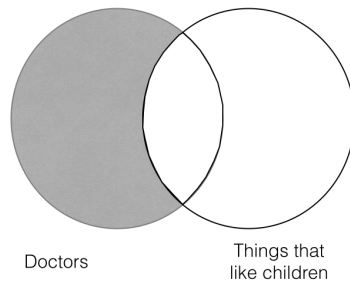
that there is nothing in the "doctors" category that is outside the

"things that like children category." However, the premises do not represent that information, since the section of the category "doctors" that lies outside of the intersection of the category "things that like children" is unshaded, thus representing that there can be things there.

Sometimes when filling in particular statements on a three category for the premises, you will encounter a problem that requires another convention in order to accurately represent the information in the Venn. Here is an example where this happens:

1. Some mammals are bears
2. Some two-legged creatures are mammals
3. Therefore, some two-legged creatures are bears

There are three categories referred to in this categorical syllogism:

Mammals

Bears
Two-legged creatures

As always, we will put the "S" term of the conclusion on the left of our three category Venn, the "P" term on the right, and the remaining term in the middle, as follows:
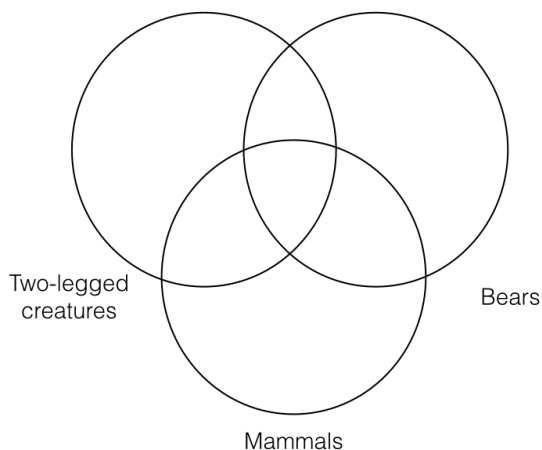
Now we need to represent the first premise, which means we need to put an asterisk in the intersection of the "mammals" and "bears" categories. However, here we have a choice to make. Since the intersection of the "bears" and "mammals" categories contains a section that is outside the "two-legged creatures" category and a section that is inside the "two-legged creatures" category, we must choose between representing the particular as part of the "two-legged creatures" category or not.
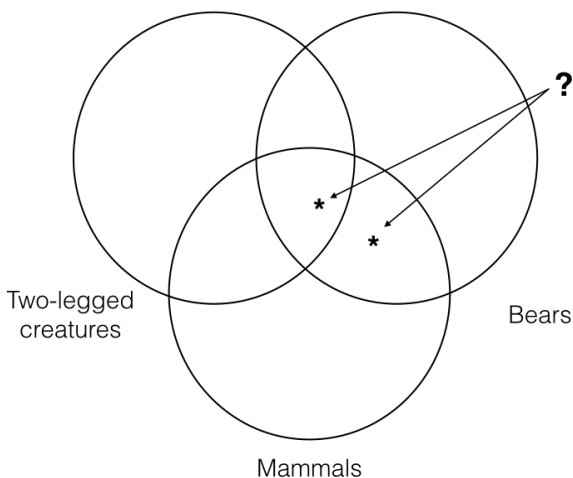
But *neither* of these can be right, since the first premise says *nothing at all* about whether the thing that is both a bear and a mammal is two-legged! Thus, in order to accurately represent the information contained in this premise, we must adopt a new convention. That convention says that when we encounter a situation where we must represent a particular on our three category Venn, but the premise says nothing about a particular category, then we must put the asterisk *on the line of that category* as I have done below. When we do this, it will represent that the particular is neither inside the category or outside the category.

Asterisk on the line of the "two-legged creatures" category

Two-legged creatures

Bears
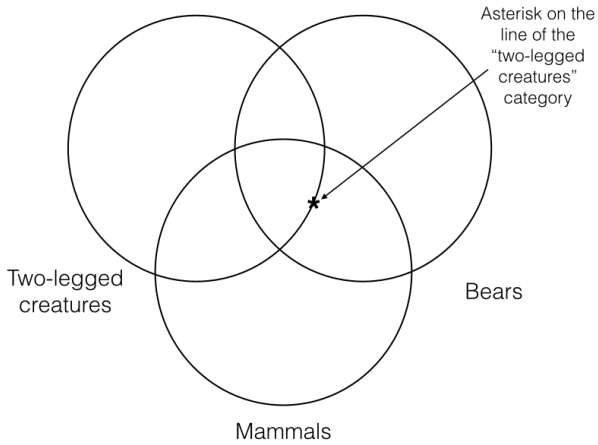
Mammals

We must do this same thing for the second premise since we encounter the same problem there. Thus, when putting the asterisk in the intersection of the "two-legged creatures" and "mammals" categories, we cannot put the asterisk either inside or outside the "bears" category. Instead, we must put the asterisk on the line of the "bears" category. Thus, using this convention, we can represent the premise Venn and conclusion Venn as follows:

Two-legged creatures

Bears

Mammals
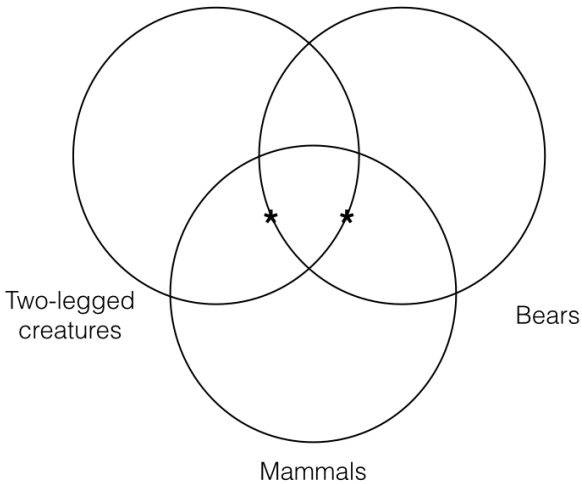
Keeping in mind the convention we have just introduced, we can see that this argument fails the Venn test of validity and is thus invalid. The reason is that the conclusion Venn clearly represents an individual in the intersection of the "two- legged creatures" and "bears" categories, whereas the premise Venn contains no such information. Thus, the conclusion Venn contains information that is not contained in the premise Venn, which means the argument is invalid.
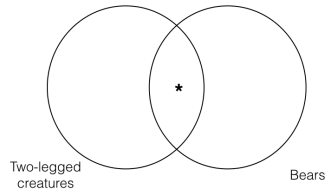
We will close this section with one last example that will illustrate an important strategy. The strategy is that we should always map universal statements before mapping particular statements. Here is a categorical syllogism that illustrates this point. This time I am going to switch to just using the capital letters S, P, and M to represent the categories. Recall that we can do this because the Venn test of validity is a formal evaluation method where we don't have to actually understand what the categories represent in the world in order to determine whether the argument is valid.

1. Some S are M
2. All M are P
3. ∴ Some S are P

If we think about mapping the first premise on our three category Venn, it seems that we will have to utilize the convention we just introduced, since the first premise is a particular categorical statement that mentions only the categories S and M and nothing about the category P:
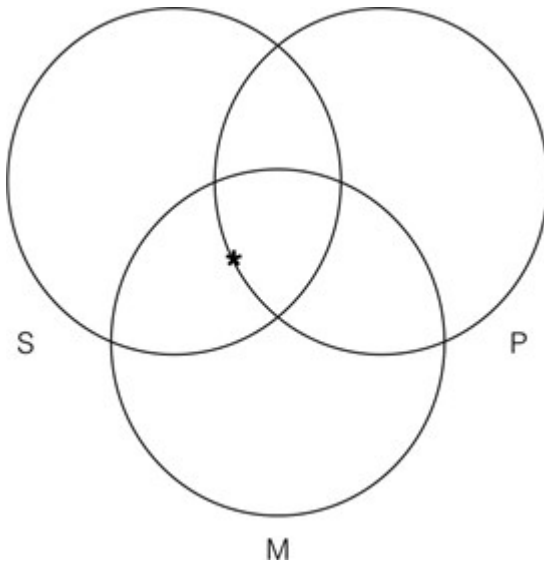
However, as it turns out, we don't have to use this convention because when we map premise 2, which is a universal statement, this clears up where the asterisk has to go:
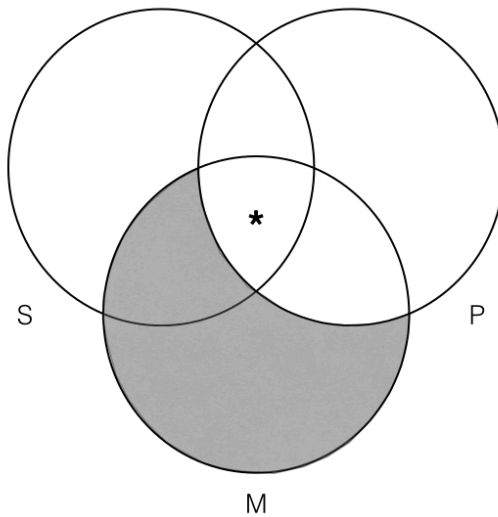
We can see that once we've mapped the universal statement onto the premise Venn (on the left), there is only one section where the

asterisk can go that is in the intersection of S and M. The reason is that once we have mapped the "all M are P" premise, and have thus shaded out any portion of the M category that is
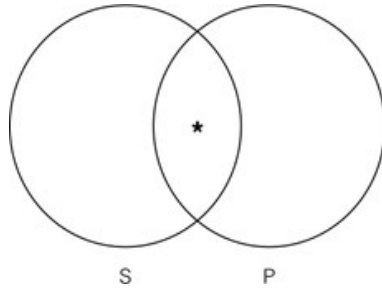
outside the P category, we know that that asterisk cannot belong inside the M category, given that it has to be inside the P category. When we apply the Venn test of validity to the above argument, we can see that it is valid since the conclusion Venn does not contain any information that isn't already contained in the premise Venn. The conclusion simply says that there is some thing that is both S and P, and that information is already represented in our premise Venn. Thus, the argument is valid. The point of strategy here is that we should always map our universal statements onto our three category Venns before mapping our particular statements. The reason is that the universal can determine how we map our particular statements (but not vice versa).

Exercise 21: Use the Venn test of validity to determine whether the following syllogisms are valid or invalid.

.   All M is P
        All M is S
        ∴ All S is P

2.   All P is M
        All M is S
        ∴ All S is P

3.   All M is P
     Some M is S
     ∴ Some S is P

4.   All P is M
     Some M is S
     ∴ Some S is P

5.   All P is M
     Some S is M
     ∴ Some S is P

6.   All P is M
     Some S is not M
     ∴ Some S is not P

7.   All M is P
     Some S is not M
     ∴ Some S is not P

8.   All M is P
     Some M is not S
     ∴ Some S is not P

9.   No M is P
     Some S is M
     ∴ Some S is not P

10. No P is M
     Some S is M
     ∴ Some S is not P

11. No P is M

Some S is not M

∴ Some S is not P

12. No M is P

Some S is not M

∴ Some S is not P

13. No P is M

Some M is not S

∴ Some S is not P

14. No P is M

No M is S

∴ No S is P

15. No P is M

All M is S

∴ No S is P

16. No P is M

All S is M

∴ No S is P

17. All P is M

No S is M

∴ No S is P

18. All M is P

No S is M

∴ No S is P

19. Some M is P

Some M is not S

∴ Some S is not P

20. Some P is M
    Some S is not M
    ∴ Some S is P

14. Attribution

The material in this chapter is adapted from previously published content:

Van Cleave, M.J. (2017). *Introduction to Logic and Critical Thinking.* Lansing Community College. Section 2.114-138, 4.185-186. https://drive.google.com/file/d/0B-rkbctcWjo3OFhGODNSYTZPMVU/ view?resourcekey=0-iGfOufS0tNl2MRQa64ITng CC BY.

EPISTEMOLOGY

15. The Rationalists

The Rationalists

Is all of our knowledge based on the evidence of the senses, or is some of it justified by other means? This epistemological question about the foundations of knowledge is what separates Rationalism and Empiricism. According to **Rationalism** at least some knowledge can be had through reason alone. For rationalists, the paradigm example of knowledge acquired independent of sense experience is mathematics. Once we have the concepts required to understand mathematical propositions (like 2+2=4), no experience is required to be justified in accepting their truth. They seem to be adequately known "through the light of reason." **Empiricism**, on the other hand, takes all of our knowledge to be ultimately grounded in sense experience. Descartes was the first significant rationalist philosopher of the modern classical period. He rejects sense experience as a trustworthy source of knowledge early in his *Meditations*. Following Descartes, a number of other European philosophers develop rationalist philosophical systems. Leibniz and Spinoza are the most notable. Meanwhile, an empiricist tradition gets started in Great Britain.

The three major empiricist philosophers are John Locke, Berkeley and David Hume. In this chapter we will focus on Descartes, Spinoza, and Liebniz, and we will take up empiricism in the next chapter.

Descartes

Rene Descartes (1596 – 1650) lived during an intellectually vibrant time. European scholars had supplemented Catholic doctrine with a tradition of Aristotle scholarship, and early scientists like Galileo and Copernicus had challenged the orthodox views of the Scholastics. Surrounded by conflicting yet seemingly authoritative views on many issues, Descartes wants to find a firm foundation on which certain knowledge can be built and doubts can be put to rest. So he proposes to question any belief he has that could possibly turn out to be false and then to methodically reason from the remaining certain foundation of beliefs with the hope of reconstructing a secure structure of knowledge where the truth of each belief is ultimately guaranteed by careful inferences from his foundation of certain beliefs.

When faith and dogma dominate the intellectual scene, "How do we know?" is something of a forbidden question. Descartes dared to ask this question while the influence of Catholic faith was still quite strong. He was apparently a sincere Catholic believer, and he thought his reason-based philosophy supported the main tenants of Catholicism. Still he roused the suspicion of religious leaders by granting reason authority in the justification of our beliefs.

Descartes is considered by many to be the founder of modern philosophy. He was also an important mathematician and he made significant contributions to the science of optics. You might have heard of Cartesian coordinates. Thank Descartes. Very few contemporary philosophers hold the philosophical views Descartes held. His significance lays in the way he broke with prior tradition and the questions he raised in doing so. Descartes frames some of the

big issues philosophers continue to work on today. Notable among these are the foundations of knowledge, the nature of mind, and the question of free will. We'll look briefly at these

three areas of influence before taking up a closer examination of Descartes' philosophy through his *Meditations of First Philosophy*.

To ask "How do we know?" is to ask for reasons that justify our belief in the things we think we know. Descartes' Meditations provide a classic example of the epistemological project of providing systematic justification for the things we take ourselves to know, and this remains a central endeavor in epistemology. This project carries with it the significant risk of finding that we lack justification for things we think we know. This is the problem of skepticism. *Skepticism* is the view that we can't know. Skepticism comes in many forms depending on just what we doubt we can know. While Descartes hoped to provide solid justification for many of his beliefs, his project of providing a rational reconstruction of knowledge fails at a key point early on. The unintended result of his epistemological project is known as the problem of Cartesian skepticism. We will explain this problem a bit later in this chapter.

Another area where Descartes has been influential is in the philosophy of mind. Descartes defends a metaphysical view known as dualism that remains popular among many religious believers. According to this view, the world is made up of two fundamentally different kinds of substance, matter and spirit (or mind). Material stuff occupies space and time and is subject to strictly deterministic laws of nature. But spiritual things, minds, are immaterial, exist eternally, and have free will. If dualism reminds you of Plato's theory of the Forms, this would not be accidental. Descartes thinks his rationalist philosophy validates Catholic doctrine and this in turn was highly influenced by Plato through St. Augustine.

The intractable problem for Descartes' dualism is that if mind and matter are so different in nature, then it is hard to see how they could interact at all. And yet when I look out the window, an image of trees and sky affects my mind. When I will to go for a walk, my material body does so under the influence of my

mind. This problem of mind-body interaction was famously and forcefully raised by one of the all too rare female philosophers of the time, princess Elisabeth of Bohemia.

A whole branch of philosophy, the philosophy of mind, is launched in the wake of problems for substance dualism. Today, the philosophy of mind is merging with neuroscience, cognitive psychology, and information science to create a new science of mind. We are rapidly learning how material brains realize the processes of thought. Once again, Descartes has failed in a most fruitful way. We also see how undeserved philosophy's reputation for failing to answer its questions is. While many distinctively philosophical issues concerning the mind remain, the credit for progress will go largely to the newly minted science of mind. The history of philosophy nicely illustrates how parenthood can be such worthwhile but thankless work. As soon as you produce something of real value, it takes credit for itself. Later in a chapter on the

philosophy of mind we will examine some developments in this area since Descartes and get acquainted with a few of its contemporary issues including the nature of consciousness.

The final big issue that Descartes brought enduring attention to is the problem of free will. We all have the subjective sense that when we choose something we have acted freely or autonomously. We think that we made a choice and we could have made a different choice. The matter was entirely up to us and independent of outside considerations. Advertisers count on us taking complete credit and responsibility for our choices even as they very effectively go about influencing our choices. Is this freedom we have a subjective sense of genuine or illusory? How could we live in a world of causes and effects and yet will and act independent of these? And what are the ramifications for personal responsibility? This is difficult nest of problems that continues to interest contemporary philosophers.

Descartes' is also a scientific revolution figure. He flourished after Galileo and Copernicus and just a generation before

Newton. The idea of the physical world operating like a clockwork mechanism according to strict physical laws is coming into vogue. Determinism is the view that all physical events are fully determined by prior causal factors in accordance with strict mechanistic natural laws. Part of Descartes' motivation for taking mind and matter to be fundamentally different substances is to grant the pervasive presence of causation in the material realm while preserving a place for free will in the realm of mind or spirit. This compromise ultimately doesn't work out so well. If every event in the material realm is causally determined by prior events and the laws of nature, this would include the motions of our physical bodies. But if these are causally determined, then there doesn't appear to be any entering wedge for our mental free will to have any influence over out bodily movements.

Now we will turn to Descartes' *Meditations* and examine how he comes to the positions just outlined. Here is a link to several of Descartes' writings including *Meditations on First Philosophy*: http://www.earlymoderntexts.com/authors/descartes

The *Meditations*

Descartes project in his meditations is to carry out a rational reconstruction of knowledge. Descartes is living during an intellectually vibrant time and he is troubled by the lack of certainty. With the Protestant Reformation challenging the doctrines of the Catholic Church, and scientific thinkers like Galileo and Copernicus applying the empirical methods Aristotle recommends to the end of challenging the scientific views handed down from Aristotle, the credibility of authority was challenged on multiple fronts. So Descartes sets out to determine what can be known with certainty without relying on any authority, and then to see what knowledge can be securely justified based on that foundation.

In the first meditation we are introduced to Descartes' method of doubt. According to this method, Descartes goes

through all of his beliefs, not individually but by categories, and asks whether there is any possible way that beliefs of this or that type can be mistaken. If so, they

must be set aside as doubtable. Many of these beliefs may ultimately be redeemed as knowledge, but they cannot serve as part of the secure foundation of indubitable beliefs from which his rational reconstruction of knowledge proceeds. Empirical beliefs, things that we believe based on the evidence of our senses, are set aside first. Our senses sometimes deceive us, as when an oar appears bent in water or a stranger in a crowd appears to be a friend. It won't do to say that we can reliably diagnose these cases and correct for mistaken appearances though because we also have experiences just like seemingly reliable sense experiences that are anything but in the case of dreams. How can we be certain that any of our seeming sense experiences of the external world aren't in fact dreams? How can we be certain that our whole life isn't a dream?

So sense experience is set to the side as uncertain and insufficient for justifying knowledge. Descartes then considers things we might know for certain by the light of reason, like mathematical claims. I seem to be about as certain in my belief that 2+2=4 as I can be about anything. Is there any possible way I could be mistaken? Descartes here imagines a powerful demon that could deceive me into always thinking that 2+2=4 when in fact this is not true. Is this a genuine possibility? Descartes allows that it is and considers all such knowledge had through reason doubtable as well.

Does anything remain? Are there any beliefs that can't be doubted, even given the hypothesis of a powerful evil deceiver? Descartes does find at least one. Even an evil deceiver could not deceive Descartes about his belief that he thinks. At least this belief is completely immune from doubt, because Descartes would have to be thinking in order for the evil deceiver to deceive him. In fact there is a larger class of beliefs about the content of one's own mind that can be defended as indubitable

even in the face of the evil deceiver hypothesis. When I look at the grey wall behind my desk I form a belief about the external world; that I am facing a grey wall. I might be wrong about this. I might be dreaming or deceived by an evil deceiver. But I also form another belief about the content of my experience. I form the belief that I am having a visual experience of greyness. This belief about the content of my sense experience may yet be indubitable. For how could the evil deceiver trick me into thinking that I am having such an experience without in fact giving me that experience? So perhaps we can identify a broader class of beliefs that are genuinely indubitable. These are our beliefs about the contents of our own mind. We couldn't be wrong about these because we have immediate access to them and not even an evil deceiver could misdirect us.

The problem Descartes faces at this point is how to justify his beliefs about the external world based on the very narrow foundation of his indubitable beliefs about the contents of his own mind. And this brings us to one of the more famous arguments in philosophy: Descartes' "Cogito Ergo Sum" or "I think, therefore I exist." Descartes argues that if he knows with certainty that he thinks, then he can know with certainty that he exists as a thinking being. Many philosophers since then have worried about the validity of this inference. Perhaps all we are entitled to infer is that there is thinking going on and we move beyond our indubitable foundation when we attribute that thinking to an existing subject (the "I" in "I exist"). There are issues to explore here. But bigger problems await Descartes, so we will just note this one and let it pass.

16. Hume

David Hume

Of the philosophers discussed here, David Hume (1711-1776) has probably had the greatest influence on contemporary analytic philosophy. The twentieth century begins with a movement known as Logical Positivism that tests the limits of Empiricism. The Empiricism of the Logical Positivists is heavily indebted to Hume.

Hume's empiricist epistemology is grounded in his philosophy of mind. Hume starts by asking what we have in the mind and where these things come from. He divides our mental representations into two categories, the relatively vivid **impressions**, these include sensations and feelings, and the less vivid **ideas** which include memories and ideas produced by the imagination.

What distinguishes impressions from ideas in our experience is just their vividness. The picture of the mind Hume offers is one where all of our beliefs and representations are cooked up out of basic ingredients provided by experience. Our experience gives us only impressions through sense experience and internal impressions like feelings. From this we generate less vivid ideas. Memories are merely faint copies of impressions. Through the imagination we can generate further ideas by recombining elements of ideas we already have. So through impressions we get the idea of a lizard and the idea of a bird. We can then generate the idea of a dragon by imaginatively combining elements of each. In cooking up new ideas from old ideas, the imagination is guided by associating relations like resemblance, contiguity (next-to-ness) and cause and effect. So, for example, an impression of a grapefruit might lead me to think of an orange due to their similarity. The thought of my

bicycle might lead me to think of the table saw it is parked next to in the basement. Through the association of cause and effect, my idea of a struck match leads me to the idea of a flame. The last of these principles of association, cause and effect, turns out to be faulty for reasons we will examine shortly.

The imagination is not merely a source of fancy and fiction. The imagination also includes our ability to understand things when we reason well in formulating new ideas from old ones. *A priori* reasoning, which is reasoning independent of experience, can produce understanding of relations of ideas. Mathematical and logical reasoning is like this. When I recognize the validity of an argument or the logic behind a mathematical proof, the understanding I attain is just a matter of grasping relations between ideas. But *a priori* reasoning only reveals logical relations between ideas. It tells us nothing about matters of fact. Our ability to understand matters of fact, say truths about the external world, depends entirely on *a posteriori* reasoning, or reasoning based on experience. As we will see, our ability to reason about matters of fact doesn't get us very far.

Often our philosophical confusion is the result of having added more than we are entitled to add to our experience when we are striving to understand it. Hume aims to correct many of these errors and, in doing so, he aims to delineate the limits of human knowledge and understanding. As it turns out, we don't know as much as we commonly suppose, in Hume's opinion. The result of Hume's rigorous Empiricism is skepticism about a great many things. Some of Hume's skeptical results are not so surprising given his Empiricism. Hume is skeptical about objective moral truths, for instance. We don't get to observe rightness and wrongness in the way we can see colors and shapes, for instance. The idea that there are objective moral truths, according to Hume, is a mistaken projection of our subjective moral sentiments.

Hume is not worried that his subjectivism about morality will lead to moral anarchy. Note that the opinion that it's OK to

do whatever you want is itself a moral opinion. So, for the subjectivist, "anything goes" is no more rationally justified than any other moral opinion. While Hume does think that morality is concerned with subjective sentiments, not objective facts, the lack of objective moral truths won't corrupt us or undermine the social order because we all have pretty much the same sorts of moral sentiments and we can base a sensible social order on these. While we may feel differently about specific practices or principles, Hume thinks we have a basis for negotiating our moral differences in our more general and more or less universally shared moral sentiments of self-love, love for others, and concern for happiness.

Hume's skepticism about objective moral truths now strikes many people as common sense. But the empiricist epistemology that leads him to subjectivism about morality also leads him to skepticism about causation, the external world, inductive reasoning, about God, and even about the self. We'll examine these further skeptical conclusions starting with causation.

Causation

When we examine our everyday idea of causation, Hume says we find four component ideas:

- the idea of a constant conjunction of cause and effect (whenever the cause occurs, the effect follows).
- the idea of the temporal priority of the cause (the cause happens first, then the effect).
- the idea of causes and effects being contiguous (next to each other) in space and time.
- the idea of a necessary connection between the cause and the effect.

So, for instance, the idea that striking a match causes it to light is made up of the idea that whenever similar matches are struck (under the right conditions), they light, plus the idea of the striking happening first, and the idea of the striking and the lighting happen right next to each other in time and space, and, finally, the idea that the striking somehow necessitates or

makes the match light. Now let's consider these component ideas and ask whether they all have an empirical basis in corresponding sense impressions. We do have sense impressions of the first three: the constant conjunction of cause and effect, the temporal priority of the cause, and the contiguity of cause and effect. But Hume argues that we lack any corresponding empirical impression of necessary connections between causes and effects. We don't observe anything like the cause making the effect occur. As Hume puts the point,

> When we look about us towards external objects, and consider the operation of causes, we are never able, in a single instance, to discover any power or necessary connexion; any quality, which binds the effect to the cause, and renders the one an infallible consequence of the other. We only find, that the one does actually, in fact, follow the other. (*An Enquiry Concerning Human Understanding*, Section VII)

The idea of causes necessitating their effects, according to Hume's analysis, is a confused projection of the imagination for which we find no basis in experience. For this reason, Hume denies that we have rational grounds for thinking that causes do necessitate their effects.

The External World

All of our reasoning about the external world is based on the idea of causation. So the skepticism that follows from Hume's skepticism about causation is quite far reaching. Our beliefs about the external world, for instance, are based on the idea that things going on in the external world cause our sense impressions. We have no rational grounds for thinking so, says Hume.

More generally, our evidence for what we can know begins with our impressions, the mental representations of sense experience. We assume that our impressions are a reliable guide to the way things are, but this is an assumption we can't rationally justify. We have no experience beyond our impressions that could rationally certify that our impressions correspond in any way to an external reality. Our assumption that our impressions do correspond to an external reality is a rationally unsupportable product of our imagination.

Induction

Closely related to Hume's skepticism about causation is Hume's skepticism about inductive reasoning. Inductive argument, in its standard form, draws a conclusion about what is generally the case, or what will prove to be the case in some as yet unobserved instance, from some limited number of specific observations. The following is an example of a typical inductive argument:

1. Every observed sample of water heated to well over 100 C has boiled.
2. Therefore, whenever water is heated to well over 100 C, it boils.

Unless every instance of water heated to over 100C in the history of the universe is among the observed instances, we can't be sure that the conclusion is true given the truth of the premises. It follows that strong inductive arguments like the one above are not deductively valid. But then what justifies the inference from the premise to the conclusion of an inductive argument?

Hume considers the suggestion that every inductive argument has a principle of induction as a suppressed premise, and it is this principle of induction that renders the inference from premises to conclusion rational. This principle of induction tells us roughly that unobserved instances follow the pattern of observed instances. So inductive arguments really go something like this:

1. Every observed sample of water heated to over 100 C has boiled.
2. <u>(Unobserved cases tend to follow the pattern of observed cases)</u>
3. So, whenever water is heated to over 100 C, it boils.

Of course the argument still isn't valid, but that's not what we are aiming for in induction. Given the hidden second premise – our principle of induction – we can reasonably hold that the premises taken together give us good grounds to accept that the conclusion is probably true. However, if this principle of induction (2 above) is to render inductive inferences rational, then we need some grounds for thinking that it is true. In considering how this principle of induction is to be justified, Hume presents a dilemma. Since there is no contradiction in denying the principle of induction, it cannot be justified *a-priori* (independent of our experience as can be done with logical truths). And any empirical argument would be inductive and therefore beg the question by appealing to the very principle of

induction that requires support. So, Hume concludes, we have no rational grounds for accepting inductive inferences.

Think about the ramifications of Hume's skepticism about induction. If inductive argument is not rational, then we have no reason at all to think the sun will rise tomorrow. Here we aren't worried about improbably possibilities like the sun getting blown to bits by aliens before tomorrow morning. Hume's argument against the rationality of inductive reasoning implies that all of our experience of the sun regularly rising gives us no reason to think its rising tomorrow is even likely to happen. If this sounds crazy, then we have a problem because it is not easy to find a defect in Hume's reasoning. This is why philosophers speak of this topic as the *Problem of Induction.* Very few are prepared to accept Hume's skepticism about induction. But in the two and a half centuries that have passed since Hume died, we have yet to settle on a satisfactory solution to the problem of induction. We'll take a closer look at this problem when we take up the Philosophy of Science in the next chapter.

God

Unlike Locke and Berkeley, Hume's rigorous Empiricism leads him to skepticism about religious matters. To avoid censorship or persecution, critics of religious belief in the 18th century exercised caution in various ways. Hume's earliest challenge to religious belief, an essay on miracles, was removed from his early work, his *Treatise of Human Nature*, and published only in his later *Enquiries Concerning Human Understanding*. In this essay, Hume argues that the belief in miracles can never be rational. A miracle is understood to be a violation of the laws of nature resulting from Divine will. But, argues Hume, the weight of the evidence of our experience overall will always give us stronger reason to mistrust our senses in the case of a

seemingly miraculous experience than to doubt the otherwise consistently regular course of events in our experience. Testimony by others of miracles is on even shakier ground.

> No testimony is sufficient to establish a miracle, unless the testimony be of such a kind, that its falsehood would be more miraculous than the fact which it endeavors to establish. (*Enquiries Concerning Human Understanding*, Section 10)

Among educated people in the 18th century, religious belief was thought to be supported not just by Divine revelation, but by our experience of the natural world as well. When we look to the natural world we find impressive harmony in the natural order of things. The various species all seem well suited to their environments and ecological stability is maintained by the various roles organism play in their environments. To the discerning mind in search of an explanation, the order and harmony we find in the world looks very much like the deliberate work of a Divine creator. This line of thought is known as the Argument from Design. Hume's last work, his posthumously published *Dialogues of Natural Religion*, aimed to undermine many arguments for the existence of God, including the Design Argument.

According to Hume, the Design Argument is a weak argument by analogy. We have reason to think that machines are the product of human design because we are familiar with their means of production. But we have no analogue in the case of the universe. We have not observed its creation. The alleged similarity of the universe to machines designed by humans is also suspect. We do find regularities in nature, but only in the small corner of nature we are familiar with. The regularity, order, and harmony we do find don't provide enough of the appearance of design to warrant positing an intelligent designer, according to Hume. But suppose we do think the natural world bears the marks of a designer's craftsmanship. The only sorts of designers

we are familiar with are people like us. But that doesn't tell us much about what sorts of being could be designers of complex harmonious systems. So even assuming we find the appearance of design in nature, we have little grounds to think that it is the product of a personal god or any sort of entity we can relate to.

Charles Darwin's theory of evolution by natural selection provides a naturalistic account of the appearance of design in life forms. Thanks to providing a developed naturalistic alternative to the hypothesis of design by a Divine creator, Darwin probably had the greater impact in undermining the design argument for the existence of God. Darwin cites Hume as among his major influences, and there are a number of passages in Hume's writing that foreshadow insights that Darwin developed.

17. Attribution

The material in this chapter is adapted from previously published content:

Payne, R.W. (2015). *An Introduction to Philosophy*. BCcampus. Section 4.39-41, 5.54-56. https://open.umn.edu/opentextbooks/textbooks/598 CC BY NC.

THEORIES OF EXISTENCE

18. Ontological Arguments for God's Existence

A quite different version of the cosmological argument is presented by William Lane Craig, drawing upon the Islamic philosophers of the 9th-12th centuries such as al-Ghazali (1058-1111), called the kalām cosmological argument. Craig argues that whatever begins to exist has a cause, that the universe began to exist, and that God must be invoked as its cause. Why believe that the universe began to exist? For one thing, it seems that the universe cannot have an infinite temporal duration since the successive addition of finites cannot add up to something infinite. Just as one cannot "count to infinity," the compounding of the moments that pass in time could not ever add up to an infinite temporal duration. For another, if we make the supposition that the universe has an infinite temporal duration various absurdities arise. Sundays are a subset (one-seventh) of all the days that have ever occurred. A very bored deity would count out six non-Sundays for every Sunday. But if the universe has an infinite temporal duration, then an infinite number of Sundays have occurred. And an infinite number of non-Sundays have occurred. Therefore, the subset is equal in magnitude to the set—an absurdity. So, the universe began to exist. Notice that Craig's argument avoids referring to necessary beings, or the principle of sufficient reason; Craig's argument requires only that if something begins to exist, then it has a cause. Supporters of the kalām cosmological argument may also cite scientific evidence to support the idea that the universe began to exist, for instance the Big Bang theory or the idea that if the universe had an infinite temporal duration, then **entropy** would guarantee that complex matter would not exist presently (Craig 1979).

One key question about Craig's kalām cosmological argument is whether the cause of the universe must be something like our

conception of God, a kind of personal agent. Craig, following al-Ghazali, suggests that the cause of the universe must be timeless, outside of time entirely. Physical causes bring about their effects, as it were, immediately. For example, an effect like the process of water freezing will begin to happen as soon as its cause, a sub-zero temperature, is present. So, if the cause of the universe is timeless and is a physical cause, we would expect the universe to have always existed. But as we have seen, that cannot be. So, the cause of the universe must be non-physical. Aside from physical causes, we sometimes explain effects as resulting from actions—we have the idea that personal agents bring about effects spontaneously as and when they will to do so, in a way that is different than and not entirely determined by physical causes. On this model, plausibly the cause of the universe is the action of a personal, but non-physical, agent. Others have objected, though, that it is difficult to make sense of the idea of a personal agent who acts but is also outside of time, and again that we are having to rely too heavily on our limited repertoire of concepts: for all we know, there might be causes that are neither like the physical nor like personal agency.

Questions to Consider

4. It seems that the opponent of the cosmological argument can try to defuse it by denying that the universe has a reason for its existence, or a cause, or by denying the principle of sufficient reason. Are these unreasonable moves? Is there any claim or principle that it would be unreasonable to deny, if the alternative was the conclusion that God exists?

5.   In theory, could science one day prove that the universe did not begin to exist? What impact would such a finding have on Clarke's cosmological argument? On Craig's kalām cosmological argument?

6.   Is it reasonable to rely on our limited repertoire of concepts, as exemplified in the discussion about whether the cause of the universe is a personal agent? Should we be worried by the thought that reality may be stranger than we can conceive?

THE ONTOLOGICAL ARGUMENT

"Ontos" being Greek for "being" or "existence," the **ontological** argument is unusual in that it has no empirical premises at all; God is not called upon as an explanation for anything. Rather, God's existence is proven by reflection on the concept of God. This is an extremely unfamiliar way of proceeding, since ordinarily we think that by **analyzing** the concept of something, we may discover the **predicates** that will be true of it *if* it exists, but not *that* it exists. For instance, if I have a child then the predicate "has a grandfather named Patrick" will be true of it. The ontological argument proposes, in the case of God, to abolish this "if" and proceed directly from the concept of God to his existence. The argument's first proponent was Anselm of Canterbury (1033-1109). It's a familiar idea that God is great, the greatest in fact, so great one cannot think of anything greater. Anselm draws on this familiar idea in his *Proslogion*. There, Anselm characterizes God as "a being than which nothing greater can be conceived" (Anselm 1078). In more modern language, Anselm is saying that God is the greatest conceivable being, that it is part of the concept of God that it is impossible to

conceive of any being greater than God. It seems that existence is greater than nonexistence. So, if we conceive of God as nonexistent, then we can conceive of something greater than God: e.g., a shoe, a flea. But God is the greatest conceivable being, so our assumption of God's nonexistence must have been false, and God must exist. Another way of putting this is that Anselm anticipates Hume's objection that no being's existence is necessary (since any being's nonexistence can be conceived without contradiction). Anselm insists that in this case the idea of God, properly understood, does give rise to contradiction if we suppose his non-existence. "The being which must exist does not exist" seems like a contradiction.

From the outset, the ontological argument has had difficulties heaped upon it. For one thing, although it may seem intuitively right that existence is greater than nonexistence, what does "greater" mean? Better than? Preferable to? More real than? A satisfying characterization is hard to find. Another early objection comes from Gaunilo of Marmoutier (994-1083), who makes the parodic suggestion of an island that is the greatest island that can be conceived. If such an island is to be greater than, say, Corsica, it must exist. Must we then say that such an island exists? Surely not. The difficulty raised by Gaunilo is that it seems that the predicate of existence can be bolted on to any concept illicitly. Anselm responds, however, that his argument applies uniquely to the greatest *being* that can be conceived (not a given, limited kind of being like an island), since although the imagined island would indeed be greater if it existed, it is not part of the concept of anything except the greatest being that can be conceived that it be greater than everything else, and so for it alone can we infer its existence from its concept. A similar response is that contingency is part of the concept of an island (or dog, or horse, or any other specific, limited kind of being which we are acquainted with), so that a necessarily existing island would simply be a contradiction. Only with the non-specific concept of "a being" in general would contingency not just be included in the concept.

The most historically influential criticism of the ontological

argument, however, comes from Immanuel Kant (1724-1804). In his *Critique of Pure Reason*, Kant argues that existence is not a predicate (Kant 1781). Think about the concept of a banana. We can attribute certain predicates to it, such as "yellowness" and "sweetness." As time goes by, we might add further predicates to the concept, e.g., "nutritional potassium source." Now think about what happens to the concept of a banana when you suppose that bananas exist. It seems that the concept is not changed at all. To say something exists is not to say anything about the concept of it, only that the concept is instantiated in reality. But if existence cannot be part of a concept, then it cannot be part of the concept of God, and cannot be found therein by any sort of analysis.

Kant's argument was widely taken to be calamitous to the ontological argument. However, in the 1960s, the argument was rejuvenated, in a form that (perhaps) avoids Kant's criticism, by Norman Malcom (1911-1990). Malcolm suggests that although existence may not be a predicate, necessary existence is a predicate. As contingent beings, we are the sort of things which can come into and go out of existence. But if God exists, then he is a necessary being rather than a contingent being. So, if he exists he cannot go out of existence. This is a predicate God enjoys, even if existence per se is not a predicate (Malcolm 1960). Intuitively, "indestructibility" and "immortality" are predicates that alter the concept of a thing. Another modern version of an Anselmian ontological argument is offered by Lynne Rudder Baker (1944-2017). Baker's version avoids the claim that existence is a predicate (as well as several other traditional difficulties). Instead, Baker notes that individuals who do not exist have mediated causal powers, that is, they cause effects but only because individuals who do exist have thoughts and beliefs about them: Santa Claus has the mediated causal power to get children to leave cookies out for him, children who themselves have unmediated causal powers. In short, to have unmediated causal powers is intuitively greater than having mediated causal powers, so given that God is the greatest being that can be conceived of, God

must have unmediated causal powers, and so he must exist (Baker 2013).

A final difficulty that we may mention for these three theistic proofs is whether they prove the existence of the God of Abraham, or the God of classical theism (supposing that the two are the same) — which it is the concern of most theistic philosophers to do. The teleological argument may show a designer, which corresponds tolerably well to the creatorhood of God, but seems to fall short of showing God's other attributes, like omnibenevolence. Similarly, the world-cause or necessary being purportedly shown by the cosmological and ontological arguments may seem far distant from a personal God who is interested in our affairs. One theistic response is that these arguments may work in combination, or be supplemented by the evidence of revelations, religious experiences, and miracles (See Chapter 3 for a few such arguments), or we may be able to find ways in which one divine attribute implies the others. Bear in mind also that there are many less well-known theistic arguments beyond these three traditional ones (McIntosh 2019). (For some specific examples, see Chapter 3.)

Questions to Consider

7. Do we really have a conception of "a being than which nothing greater can be conceived"? Is that something we are able to frame in our minds, or have we just begun to misuse words?

8. If existence is not a predicate, why do we treat it as one in ordinary sentences, like "the pecan tree exists"? Further, how do we delineate the domain of fiction? Isn't our concept of "Homer Simpson" a

concept of a character who does not exist? If not, what is it a concept of?

9.   Even once you grasp it, does the ontological argument seem intuitive to you? Does it seem less intuitive than the cosmological argument? Should you put much weight on your intuitions about these arguments?

19. Reasons to Believe – Theoretical Arguments

Thinking about God brings together our powers of speculation, our deepest values, and our greatest hopes and fears. It is therefore fertile philosophical territory. Some of the arguments for belief in God are theoretical in that they appeal to our reason. Other arguments are practical in that they invoke God to make sense of some of our practices, such as morality. In this chapter, we will review the most influential theoretical arguments for God's existence: the teleological, the cosmological, and the ontological arguments. The former two try to show God's existence using tools familiar from ordinary empirical reasoning; God is a hypothesis to be proven in much the same way as we prove more mundane hypotheses, marshalling the evidence as best we can. Just as a one might see a puddle and infer that it has been raining recently, one might observe certain other features of the world and infer God as the best (or only) explanation of them. The latter argument is more closely akin to mathematics and conceptual analysis; just as one might reflect on the concept of a triangle and ascertain that its internal angles must add up to 180°, one might reflect on the concept of God and ascertain that he must exist. Lastly, we will introduce the suggestion that it is legitimate to believe in God without providing arguments at all: that belief in God is more properly a cornerstone for our thinking, than a mere conclusion of some argument. Each of these arguments have been articulated in myriad ways, so we will focus our attention on some of the most influential versions.

The Teleological Argument

"Telos" being Greek for "purpose" or "goal," the **teleological** argument takes as its starting point the appearance of purpose or design in the world. If there is design, there must be a designer. This thought is an ancient and cross-cultural one, appearing in classical Hindu thought (Brown 2008) and in the Psalms: "The heavens declare the glory of the Lord; and the firmament sheweth his handiwork" (Psalm 19:1). An influential formulation comes from William Paley (1743-1805). In *Natural Theology*, Paley offers numerous instances of apparent design, focusing primarily on biological organisms. Paley argues that organisms are analogous to human-created artifacts in that they involve a complex arrangement of parts that serve some useful function, where even slight alterations in the complex arrangement would mean that the useful function was no longer served. An eye, like a watch, evidently serves a useful function. The function is only achieved by a very complex arrangement of parts, which in turn serve various sub-functions, all ordered towards the higher function. Had this arrangement been different in any minute detail, the eye would not successfully serve its higher function. To explain this feature of the eye, we should, on an **analogy** with the watch, refer to a designing mind's activity, rather than the blind play of causal forces. As we are to the watch, so God is to the eye. To Paley, God is a powerful and simple hypothesis that must be invoked to explain the design resplendent in nature (Paley 1802).

Formulations of the teleological argument like Paley's have been subjected to searching criticisms, not least by David Hume (1711-1776). In his fabulously written *Dialogues Concerning Natural Religion*, Hume questions how close the analogy of design really is. For example, we produce artifacts by acting on pre-existing materials, but God is supposed to create from nothing. Most artifacts have a purpose that is evident to us, but God's purpose in having created this or that creature, or the world at all, is unclear.

We have seen artifacts being manufactured on many occasions, but never an organism, or the world. Even granting unequivocally that there is design in the world, we would not be justified in inferring God to explain it. Hume notes that artifacts are usually the result of collaboration by many people. Nor is there any connection between the qualities of an artifact and the qualities of its designer; one need not be a giant to build a skyscraper or be beautiful to make a beautiful painting. So, the design in the world need not be the design of one being, or an especially exalted being. Rather, the evidence of design is equally consistent with the hypothesis of polytheism (Hume 1779). Perhaps as devastating for Paley's formulation, Charles Darwin's (1809-1882) theory of evolution by natural selection is widely taken to show that the complex arrangement of parts and the functions of the parts of organisms can be accounted for without reference to a designing mind. The appearance of design is merely appearance; the analogy between artifacts and organisms is a misleading one. God is an obsolete hypothesis so far as the explanation of these phenomena are concerned. A distinct minority, the proponents of "Intelligent Design" contest this claim by offering examples of biological phenomena that supposedly cannot be explained by Darwinian evolution (Behe 1996). Barbara Forest argues that "Intelligent Design" theories lack a serious methodology, given that they invoke miraculous intervention in an unprincipled way to explain various phenomena (Forrest 2011).

However, teleological arguments continue to thrive in other forms. One line of thinking is the fine-tuning argument. Our universe seems to be governed by a batch of laws of nature—e.g. gravity, the strong nuclear force. It seems possible that these laws of nature could have been different in an unfathomable number of ways—e.g. we can conceive gravity as a billion times stronger than it is, or a billion times weaker. It seems that most of the ways that the laws of nature could have been would not allow for embodied moral agents (or, more broadly, life) by not allowing for the emergence of complex matter. Now, arguably God is a being

who wishes there to be embodied moral agents. So, if there is a God, this predicts a universe with laws of nature that allow for the emergence of embodied moral agents, laws that are finely-tuned for such a purpose. By contrast, if there is no God there is no particular reason to predict that the laws of nature will be like this. Our universe seems to be one with laws that allow for embodied moral agents. Therefore, our universe is more consistent with the theistic hypothesis, so probably God exists. Finally, putting aside the fine-tuning of the physical laws we enjoy, Richard Swinburne contends that the fact that our universe is governed by laws at all, rather than being chaotic, is something that demands a design-based explanation (Swinburne 2004).

Whether such arguments really identify phenomena that stand in need of a special explanation, and whether the explanations they offer are vulnerable to being supplanted by non-theistic alternatives, is a matter of ongoing debate.

Questions to Consider

1. What is the value of arguments by analogy, such as Paley's? Do they give new information, or just highlight information you already had, or can they even be misleading?
2. Suppose you were convinced that our universe is in fact fine-tuned. What, if anything, would you be entitled to infer about the nature of the fine-tuner(s)?
3. Many have thought that Darwinian evolution thoroughly undermines the view that biological phenomena are designed by God. Is there a consistent way of holding both views? Supposing

there is, would the hypothesis of a designer-God still be a necessary part of the explanation of the biological phenomena, or a somewhat ornamental addition?

The Cosmological Argument

"Cosmos" being Greek for "world," the **cosmological** argument suggests God as the only adequate hypothesis in explaining why there is something rather than nothing. Cosmological arguments go back at least as far as Plato (428-348 BCE), with influential formulations being offered by Thomas Aquinas (1225-1274) and Gottfried Leibniz (1646-1716). One influential formulation comes from Samuel Clarke (1675-1729).

In *A Demonstration of the Being and Attributes of God*, Clarke argues for the conclusion that God is the reason for the universe's existence by showing the bankruptcy of the alternatives. Something must have existed from eternity, Clarke reasons, since to suppose otherwise would be to suppose that something arose from nothing, which is absurd. Further, this eternal something must be independent of the universe. Think of a sapling tree. Like every individual thing in the universe its existence is **contingent**—it could fail to exist—as demonstrated by the fact that it once did not exist and by the fact that it is susceptible to change and destruction. Therefore, its reason for existing must be sought outside it; if we seek the reason why the sapling exists we must refer to its parent tree, the soil, the sun, the air. But if everything in the universe is contingent, then so is the universe itself, and its reason for existing must be sought outside it. Even if the universe had no beginning in time, and we could trace the sapling's reason for existing backward

indefinitely, we would still need to explain why there was this endless succession of contingent beings rather than nothing. Think of "reason for existing" as being like the parcel in the children's game "pass the parcel."[1] Even supposing an infinite number of players, or a circle of players passing the parcel for an eternity, if every player must receive the parcel from another (like a contingent being receives its reason for existing from another), then we would still face the question where the players got the parcel in the first place. Lastly, the being outside the universe must have a necessary existence; that is, it must contain the reason for its existence within itself, such that it could not fail to exist. By the difficulties attending all the alternatives, we are driven to accept that not all beings are contingent; our search for reasons for existing must reach its terminus in a necessary being, God. Clarke admits that the notion of necessary existence is difficult to conceive, since all the beings we encounter are contingent, but holds that it is the only adequate hypothesis in explaining why there is something (Clarke 1705).

Clarke's cosmological argument was also criticized by Hume in his *Dialogues Concerning Natural Religion*. Hume questions why the universe itself may not be the necessary being. Clarke's reason for rejecting this idea was that everything in the universe is contingent. But, Hume notes, Clarke is committing the fallacy of composition. A flock may be composed of sheep destined for slaughter, but this does not prove that the flock itself is destined for slaughter. Likewise, perhaps the universe's existence is necessary despite the contingency of every individual thing in it, a thought which is lent some credibility by the physical principle that matter can neither be created nor destroyed. Raising further havoc, Hume questions whether there can even be such a thing as a necessary being. It seems to be a feature of claims which are necessary—like "2+2=4" or "a nephrologist is a physician of the kidneys"—that their contraries cannot be conceived without contradiction, as with "2+2=5." But we seem able to conceive any being's nonexistence without contradiction; just as I can coherently conceive of the sapling's

nonexistence, I can coherently conceive of God's nonexistence (as shown by the fact that we feel the need to debate God's existence).

Another issue is that Clarke's cosmological argument, like many other formulations, invokes the "principle of sufficient reason," or the idea that every state of affairs has a reason why it is so and not otherwise. This seems to be a principle that we make thorough use of from early childhood in endlessly asking "why?" and expecting that there must be answers. Because of this principle, we insist that the universe must have a reason for its existence, rather than allowing that the universe is an unaccountable "brute fact." But why should we accept the principle of sufficient reason? It does not seem to be a necessary truth or something we can infer from experience (Pruss 2006).

A quite different version of the cosmological argument is presented by William Lane Craig, drawing upon the Islamic philosophers of the 9th-12th centuries such as al-Ghazali (1058-1111), called the kalām cosmological argument. Craig argues that whatever begins to exist has a cause, that the universe began to exist, and that God must be invoked as its cause. Why believe that the universe began to exist? For one thing, it seems that the universe cannot have an infinite temporal duration since the successive addition of finites cannot add up to something infinite. Just as one cannot "count to infinity," the compounding of the moments that pass in time could not ever add up to an infinite temporal duration. For another, if we make the supposition that the universe has an infinite temporal duration various absurdities arise. Sundays are a subset (one-seventh) of all the days that have ever occurred. A very bored deity would count out six non-Sundays for every Sunday. But if the universe has an infinite temporal duration, then an infinite number of Sundays have occurred. And an infinite number of non-Sundays have occurred. Therefore, the subset is equal in magnitude to the set—an absurdity. So, the universe began to exist. Notice that Craig's argument avoids referring to necessary beings, or the principle of sufficient reason; Craig's argument requires only that if something begins to exist, then it has a cause. Supporters of the kalām

cosmological argument may also cite scientific evidence to support the idea that the universe began to exist, for instance the Big Bang theory or the idea that if the universe had an infinite temporal duration, then **entropy** would guarantee that complex matter would not exist presently (Craig 1979).

One key question about Craig's kalām cosmological argument is whether the cause of the universe must be something like our conception of God, a kind of personal agent. Craig, following al-Ghazali, suggests that the cause of the universe must be timeless, outside of time entirely. Physical causes bring about their effects, as it were, immediately. For example, an effect like the process of water freezing will begin to happen as soon as its cause, a sub-zero temperature, is present. So, if the cause of the universe is timeless and is a physical cause, we would expect the universe to have always existed. But as we have seen, that cannot be. So, the cause of the universe must be non-physical. Aside from physical causes, we sometimes explain effects as resulting from actions—we have the idea that personal agents bring about effects spontaneously as and when they will to do so, in a way that is different than and not entirely determined by physical causes. On this model, plausibly the cause of the universe is the action of a personal, but non-physical, agent. Others have objected, though, that it is difficult to make sense of the idea of a personal agent who acts but is also outside of time, and again that we are having to rely too heavily on our limited repertoire of concepts: for all we know, there might be causes that are neither like the physical nor like personal agency.

Questions to Consider

4.   It seems that the opponent of the cosmological argument can try to defuse it by denying that the universe has a reason for its existence, or a cause, or by denying the principle of sufficient reason. Are these unreasonable moves? Is there any claim or principle that it would be unreasonable to deny, if the alternative was the conclusion that God exists?

5.   In theory, could science one day prove that the universe did not begin to exist? What impact would such a finding have on Clarke's cosmological argument? On Craig's kalām cosmological argument?

6.   Is it reasonable to rely on our limited repertoire of concepts, as exemplified in the discussion about whether the cause of the universe is a personal agent? Should we be worried by the thought that reality may be stranger than we can conceive?

THE ONTOLOGICAL ARGUMENT

"Ontos" being Greek for "being" or "existence," the **ontological** argument is unusual in that it has no empirical premises at all; God is not called upon as an explanation for anything. Rather, God's existence is proven by reflection on the concept of God. This is an extremely unfamiliar way of proceeding, since ordinarily we think that by **analyzing** the concept of something, we may discover the **predicates** that will be true of it *if* it exists, but not *that* it exists. For instance, if I have a child

Chapter Notes

1. Pass the parcel is a parlour game in which a parcel containing a prize is passed around and around in a circle.

20. Humean Arguments

Another problem arises when we question whether the omni-properties are consistent or coherent with one another. One could claim that any of the traits mentioned above is internally consistent and non-paradoxical, but that the set of traits attributed to God generates contradictions and cannot therefore be possessed by a single entity. Consider the following premise:

Omniscience interferes with free will.

If we take omniscience to include infallible knowledge of every future event, then God knows with absolute certainty that they will do x at a given time t.[5] If this is true, then it looks as though omniscience interferes with free will. But if omniscience interferes with free will, then it looks as though omniscience also interferes with omnipotence. If God cannot be mistaken about how they will act at t, then God is incapable of doing anything other than x. Thus, we arrive at:

If God lacks free will, then God lacks omnipotence.

And omniscience may also conflict with omnibenevolence. The freedom to do otherwise is often thought of as a precondition for morally good action (I am not performing a praiseworthy action if a mind control device forces me to rescue a drowning child). Yet if God infallibly knows how they will act and thus cannot act otherwise, then one could plausibly argue that there seems to be a similar lack of moral freedom with respect to their actions. So it appears as though omnibenevolence is inconsistent with omniscience, and we can add the following premise to the argument:

3. If God lacks free will, then God lacks omnibenevolence.

If these premises are all true, omniscience interferes with free will, and as a result it interferes with *both* omnipotence *and*

omnibenevolence. The argument would thus reach the following conclusion:

4. If God is omniscient, God cannot be omnipotent (2) or omnibenevolent (3).

And notice that one could present a different argument that begins with either omnibenevolence or omnipotence, and goes on to claim that either of these properties is inconsistent with the others. Consider:

1*. Omnibenevolence seems to interfere with free will
2. If God lacks free will, then God lacks omnipotence.

If omnibenevolence amounts to moral perfection, then we can infer that God necessarily does what is morally best in any given scenario. But this is just to say that God cannot do anything that is morally suboptimal. God cannot, therefore, be omnipotent if we take omnipotence to mean an ability to perform morally imperfect actions.

So it appears as though all of the omni-properties can be brought into **_prima facie_** conflict (that is, into conflict at first glance) with any of the others. If any of these inconsistencies hold water, then once again, the omniGod cannot exist, because in order to exist, they must possess a set of traits that are logically inconsistent with one another.

Questions to Consider

1.   Do you think that God can suspend the laws of logic and bring about contradictions? Why or why

not?

2.   Select one of the apparent inconsistencies between two omni-properties and respond to that apparent inconsistency on the omniGod theist's behalf.

3.   Is it open to the theist to abandon one or more omni-properties altogether? Can you think of reasons for them not to do so?

Problems of Evil

The omni-properties may be inconsistent not only with each other, but with observable or indispensable facts about the world. In this subsection we shall look at the apparent inconsistency between the omni-properties and the existence of evil. Take the following example:

Suppose in some distant forest lightning strikes a dead tree, resulting in a forest fire. In the fire a fawn is trapped, horribly burned, and lies in terrible agony for several days before death relieves its suffering. (Rowe 1979, 337)

For many philosophers, and many reflective non-philosophers, it is difficult to reconcile the existence of such evils in the world with belief in an omniGod. How could an almighty creator, who brims with loving-kindness, allow any evil to exist in the world, let alone evils of the scale and severity we see in the world today? This apparent tension between the existence of evil and the existence of the omniGod has birthed a number of arguments from evil, designed to show that belief in God is at best unreasonable and at worst outright irrational. Here, we shall focus on **moral evils**, evils for which some agent is morally responsible or blameworthy. As we

shall see at the end of this section, other evils must also be dealt with.

Of those arguments, J. L. Mackie's argument from evil has been by far the most influential. Mackie argued that belief in the omniGod is irrational because evil could not coexist with a God who possesses two of the omni-properties above. On Mackie's view, the inconsistency emerges once we begin to flesh out each of omnipotence and omnibenevolence:

1. If God is omnipotent, there are "no limits to what [they] can do" (Mackie 1955, 201).
2. If God is omnibenevolent, they are "opposed to evil, in such a way that [they] always eliminate[ ] evil as far as [they] can" (Mackie 1955, 201).

Together, premises (1) and (2) suggest that if the omniGod existed, evil would not.[6] The omniGod of Abrahamic theology is perfectly able and entirely willing to eliminate all of the world's troubles. But it is quite clear, Mackie insists, that evil does exist. The upshot of Mackie's argument, then, is that if evil exists (and it certainly seems to) then God is either not omnipotent or not perfectly good. In other words, the omniGod does not exist. David Hume articulates this position more forcefully in an oft-quoted passage from his *Dialogues Concerning Natural Religion* (Hume 1948): "is [God] willing to prevent evil, but not able? Then he is impotent. Is he able, but not willing? Then he is malevolent. Is he both able and willing? Whence then is evil?"[7]

One of the most renowned responses to such problems of evil, defended by philosophers like Plantinga (1974), is known as the free will defence. The free will defence begins with an intuitively plausible premise: free will is very valuable and ought to be preserved. More specifically, the free will defence begins by noting the import of libertarian free will, a capacity to choose your own actions without being caused to act by anything external (e.g. a mind control device or being held at gunpoint). A person exercises

libertarian free will whenever their actions are not brought about by outside interference. But this sort of free will therefore requires God's non-interference. God cannot force us to act in certain ways without thereby sacrificing libertarian free will. So they cannot coerce us into morally upstanding actions without eliminating something of great value. The crux of the free will defence is thus a dilemma. God must choose either to allow us our libertarian free will and in doing so run the risk that we will sometimes act reprehensibly, or to intercede in human life, preventing us from causing evil, but at the cost of our libertarian free will.[8] Despite possessing the omni-properties, God is faced with forced choices in much the same way we are, and it is better (or more modestly, it could be better for all we know) that God leaves our free will intact.

Many theists find this response satisfying, and it is certainly an elegant solution. But it is a solution which resolves only part of the problem. The free will defence makes sense of evils like murder and theft, which are freely chosen. But some evils seem to have nothing to do with free will at all. More specifically, some philosophers have argued that the free will defence cannot explain **natural evils**, evils for which no agent is morally responsible or blameworthy—like volcanic eruptions, forest fires, and tsunamis. How, after all, can Rowe's example above be explained by reference to free will? There is no discernible libertarian free will on which to lay blame there, since such evils are caused by natural processes. So we might think that the free will defence yields only a partial solution to the problem of evil, and that there are other cases of evil which require other solutions.

Chapter Notes

5. Note that this problem does not necessarily threaten classical theists, since on their view God is timeless.

6. Many philosophers go on to add a third premise, taking it to be a hidden or necessary premise in Mackie's argument:

3. *If God is omniscient, he knows about all of the world's evils and how to eradicate them;*

This makes the conclusion a *trilemma* instead of a *dilemma*, but the conclusion remains the same – the omniGod still does not exist.

7. Classical theists like Aquinas do acknowledge the challenge evil poses, but the argument plays out rather differently if God is immutable and impassible.

8. The argument thus assumes that God could not have created a world in which people both possess libertarian free will and never bring about evil—a questionable assumption, to be sure, but one we shall not challenge here.

21. External World

Causation

When we examine our everyday idea of causation, Hume says we find four component ideas:

- the idea of a constant conjunction of cause and effect (whenever the cause occurs, the effect follows).
- the idea of the temporal priority of the cause (the cause happens first, then the effect).
- the idea of causes and effects being contiguous (next to each other) in space and time.
- the idea of a necessary connection between the cause and the effect.

> So, for instance, the idea that striking a match causes it to light is made up of the idea that whenever similar matches are struck (under the right conditions), they light, plus the idea of the striking happening first, and the idea of the striking and the lighting happen right next to each other in time and space, and, finally, the idea that the striking somehow necessitates or makes the match light. Now let's consider these component ideas and ask whether they all have an empirical basis in corresponding sense impressions. We do have sense impressions of the first three: the constant conjunction of cause and effect, the temporal priority of the cause, and the contiguity of cause and effect. But Hume argues that we lack any corresponding empirical impression of necessary connections between causes and effects. We don't observe

anything like the cause making the effect occur. As Hume puts the point,

> When we look about us towards external objects, and consider the operation of causes, we are never able, in a single instance, to discover any power or necessary connexion; any quality, which binds the effect to the cause, and renders the one an infallible consequence of the other. We only find, that the one does actually, in fact, follow the other. (An Enquiry Concerning Human Understanding, Section VII)

The idea of causes necessitating their effects, according to Hume's analysis, is a confused projection of the imagination for which we find no basis in experience. For this reason, Hume denies that we have rational grounds for thinking that causes do necessitate their effects.

The External World

All of our reasoning about the external world is based on the idea of causation. So the skepticism that follows from Hume's skepticism about causation is quite far reaching. Our beliefs about the external world, for instance, are based on the idea that things going on in the external world cause our sense impressions. We have no rational grounds for thinking so, says Hume.

More generally, our evidence for what we can know begins with our impressions, the mental representations of sense experience. We assume that our impressions are a reliable guide to the way things are, but this is an assumption we can't rationally justify. We have no experience beyond our impressions that could rationally certify that our impressions correspond in any way to an external reality. Our assumption that our

impressions do correspond to an external reality is a rationally unsupportable product of our imagination.

Induction

Closely related to Hume's skepticism about causation is Hume's skepticism about inductive reasoning. Inductive argument, in its standard form, draws a conclusion about what is generally the case, or what will prove to be the case in some as yet unobserved instance, from some limited number of specific observations. The following is an example of a typical inductive argument:

- Every observed sample of water heated to well over 100 C has boiled.
- Therefore, whenever water is heated to well over 100 C, it boils.

Unless every instance of water heated to over 100C in the history of the universe is among the observed instances, we can't be sure that the conclusion is true given the truth of the premises. It follows that strong inductive arguments like the one above are not deductively valid. But then what justifies the inference from the premise to the conclusion of an inductive argument?

22. Attribution

The material in this chapter is adapted from previously published content:

Hunt, M.W. (2020). Reasons to believe-Theoretical Arguements. In *Introduction to Philosophy: Philosophy of Religion* (pp. 18-24). Rebus Community. https://press.rebus.community/intro-to-phil-of-religion/ CC BY.

Payne, R.W. (2015). *An Introduction to Philosophy. Section 5.54.* BCcampus. https://open.umn.edu/opentextbooks/textbooks/598 CC BY NC.

Steyl, S. (2020). Reasons not to believe. In *Introduction to Philosophy: Philosophy of Religion* (pp. 52-54). Rebus Community. https://press.rebus.community/intro-to-phil-of-religion/ CC BY.

MIND AND BODY

23. Mind-Body Problem

Second (and related), qualia are arguably *ineffable*; that is, they cannot neatly be put into words. Imagine trying to explain to a person who is blind what red looks like, or (a less extreme example) conveying to a lifelong vegetarian what tuna tastes like. While in both cases, we might attempt to use metaphors ("red is like a trumpet") to convey the character of the experience, our attempts to do so will inevitably fail to do justice to the relevant sensation.

A final alleged property is that qualia are immediately and fully apprehensible to us just by experiencing them. In this respect, they are distinct from the *objects* of our experience. Imagine that you are lying in bed at night and hear a soft thud. You may well wonder what the noise was: a falling object, a door slamming in the wind, or perhaps your housemate returning home. What you don't have to speculate about, however, is what the noise *sounded* like to you. This is something you grasped simply by hearing it. More strongly and more controversially, some philosophers have suggested that we can never make errors of judgment about our qualia. If I say something feels painful to me, for example, then it is nonsensical to suggest I might be in error.

Qualia and the Mind-Body Problem

One reason qualia have so fascinated philosophers is that they are arguably hard to explain in standard scientific terms. Many of us have probably heard neuroscientists talking about things like synapses, neurons, and different regions of the brain. It is perhaps not too difficult to see how this kind of scientific approach might explain various aspects of our behavior. We might understand perception, for example, in terms of the transmission of information from the sense organs through various processing areas of the

brain, or unusual aggression in terms of the release of some hormone or neurotransmitter. It is much harder to see, however, how these kinds of scientific descriptions could ever give us a satisfying explanation of why red looks the specific way that it does, or why cinnamon tastes like *this* and vanilla like *that*.

The challenge here is not merely to explain the neuroscience of how vision works or how our tongue relates flavour information to the brain. Important progress is being made every day in understanding questions like these, although the science still has a long way to go. Instead, the real difficulty is that while science tells us about how the brain *works*, it seems unable to tell us what experiences are actually *like*. To get an idea of the problem, imagine a person who has been completely deaf since birth who wants to know what Beethoven sounds like. Even if we had perfect brain-scanners and could show them exactly what happens to someone's neurons when they listen to music, it does not seem like this could ever properly convey to them the subjective experience of hearing the opening bars of the Choral Symphony.

This creates an apparent challenge for a scientific worldview. If science cannot fully explain qualia, then does it follow that science can only offer us a partial understanding of the universe? More strongly, one might wonder whether the seeming inexplicability of qualia in scientific terms shows that the universe we inhabit does not consist solely of things like atoms, molecules, forces, and other objects from the domain of science, but also contains distinctive, irreducibly *mental* phenomena.

The challenge is well illustrated by a famous thought experiment called "Mary's Room" developed by philosopher Frank Jackson (1982).[1] Imagine a woman called Mary who is a brilliant scientist. Specifically, we are told that she knows all the physical facts about color perception: she knows all about the physics of light, the biology of the eye, and the neuroscience of color processing in the brain.

Chapter Notes

1. Mary's Room is also discussed in Chapter 4.

24. Substance Dualism in Descartes

Introduction

René Descartes (1596-1650) was a French philosopher who is often studied as the first great philosopher in the era of "modern philosophy." He is the most famous proponent of a view called "substance dualism," which states that the mind and the body are two different substances. While the body is material (corporeal), the mind is immaterial (incorporeal). This view leaves room for human souls, which are usually understood as immaterial. Descartes argued on the basis of the Christian views that souls are immaterial and can exist separate from the body, but he emphasized that the mind alone is immaterial, whereas the other traditional functions of the souls can be explained as corporeal operations. His view and arguments were so influential that after him many philosophers referred to substance dualism under Descartes' name as "Cartesian dualism." In his explanation of the mind, the soul, and the ability of humans to understand the world around them through the powers of their minds, Descartes remains one of the most influential figures not just in modern philosophy, but throughout the history of philosophy. Even in the contemporary era, philosophers such as Gilbert Ryle (1900-1976) found worth in writing about and arguing against Descartes' views to set up their own theories. Ryle questioned whether the mind and body are in fact distinct and argued that they would not communicate with each other if they were. Ryle states:

> Body and mind are ordinarily harnessed together....[T]he things and events which belong to the physical world...are

external, while the workings of [a person's] own mind are internal....[This results in the] partly metaphorical representation of the bifurcation of a person's two lives. (1945, 11-16)

Ryle stated that, if Descartes' theory were correct, the mind would be a mere "ghost in a machine," inactive and unable to cause actions in the body (the machine). Ryle did not term Descartes' theory "substance dualism" but "Descartes' myth." Descartes' arguments for substance dualism and the immaterial nature of the mind and soul are therefore paramount to any investigation of the philosophy of mind, and are still being debated in present-day theories. On the other hand, with his interpretation of what he calls passions (most operations of a living body), he also provides incentives for a non-dualistic physicalism of the mind.

25. Property Dualism

Substances and Properties

The notion of a substance has a long history going back to Ancient Greek metaphysics, most prominently to Aristotle, and it has been understood in various ways since then. For present purposes we can say that a substance can be understood as a unified fundamental kind of entity—e.g. a person, or an animal—that can be the bearer of properties. In fact, the etymology of the Latin word *substantia* is that which lies below, that which exists underneath something else. So, for instance, a zebra can be a substance, which has properties, like a certain color, or a certain number of stripes. But the zebra is independent of its properties; it will continue to exist even if the properties were to change (and, according to some views, even if they ceased to exist altogether).

According to Cartesian dualism there are two kinds of substance: the material substance, which is extended in space and is divisible, and mental substances whose characteristic is thought. So each person is made up of these two substances—matter and mind—that are entirely different in kind and can exist independently of each other. Talking of the mind in terms of substances gives rise to a number of problems (see Chapter 1). To avoid these problems, property dualism argues that mentality should be understood in terms of properties, rather than substances: instead of saying that there are certain kinds of things that are minds, we say that to have a mind is to have certain properties. Properties are characteristics of things; properties are attributed to, and possessed by, substances. So according to property dualism there are different kinds of properties that pertain to the only kind of substance, the material substance: there are physical properties like having a

certain color or shape, and there are mental properties like having certain beliefs, desires and perceptions.

Property dualism is contrasted with substance dualism since it posits only one kind of substance, but it is also contrasted with ontological monist views, such as materialism or idealism, according to which everything that exists (including properties) is of one kind. Usually, property dualism is put forward as an alternative to reductive physicalism (the type identity theory) – the view that all properties in the world can, in principle at least, be reduced to, or identified with, physical properties (Chapter 2).

Hilary Putnam's (1926-2016) multiple realization argument is a main reason why reductive physicalism is rejected by some philosophers, and it provides an argument for property dualism. Although this argument was originally used as an argument for functionalism, since it challenges the identity of mental states with physical states, it was taken up by non-reductive physicalists and property dualists alike. According to the multiple realization argument then, it is implausible to identify a certain kind of mental state, like pain, with a certain type of physical state since mental states might be implemented ("realized") in creatures (or even non-biological systems) that have a very different physical make up than our own. For instance, an octopus or an alien may very well feel pain but pain might be realized differently in their brains than it is in ours. So it seems that mental states can be "multiply realizable." This is incompatible with the idea that pain is strictly identical with one physical property, as the identity theory seems to claim. If this is correct, and there is no possibility of reduction of types of mental states to types of physical states, then mental properties and physical properties are distinct, which means that there are two different kinds of properties in the world and, therefore, property dualism is true.

In addition to the multiple realization argument, probably the most famous argument for property dualism is the knowledge argument put forward by Frank Jackson (1982). This argument involves the imaginary example of Mary, a brilliant neuroscientist

who was raised in a black and white room. She knows everything there is to know about the physical facts about vision but she has never seen red (or any color for that matter). One day Mary leaves the black and white room sees a red tomato. Jackson claims that Mary learns something new upon seeing the red tomato—she learns what red looks like. Therefore, there must be more to learn about the world than just physical facts, and there are more properties in the world than just physical properties.

26. Functionalism

Introduction: Two Monsters We Must Avoid

While passing through the Strait of Messina, between mainland Italy and the isle of Sicily, Homer has Odysseus come upon two monsters, Scylla and Charybdis, one on either side of the strait. If Odysseus is to pass through the strait, he must choose between two very unhappy options; for if he averts one along the way, he will move in the other's monstrous reach. On the one side is roaring Charybdis, who would surely blot out—as if by colossal whirlpool—Odysseus's entire ship. (Have you ever been faced with an option so bad that you cannot believe you have to seriously consider it? Well, this is Odysseus's bleak situation.) On the other side of the strait, things fare little better for Odysseus and his war-weary crew: we have vicious Scylla, who only by comparison to Charybdis, looks like the right choice. The ship makes it through, Homer tells us, minus those who were snatched from the ship's deck and eaten alive. Six are taken, we are told, one for each of Scylla's heads. By comparison only, indeed.

   In this chapter we consider the theory of mind known as functionalism, the view that minds are really functional systems like the computing systems we rely on every day, only much more complex. The functionalist claims to sail a middle path between materialism (discussed in Chapter 2), or the joint thesis that minds are brains and mental states are brain states, and behaviourism (also discussed in Chapter 2), or the thesis that mental states are behavioural states or dispositions to behave in certain ways.

Avoiding Materialism

One the one side we have materialism, which we must avoid because there appears to be no strict identity between mental states and brain states. Even though human Freya is different than a wild rabbit in many interesting ways, we think they can both be in physical pain. Suppose that while restringing her guitar, Freya lodges a rogue metal splinter off the D string in the top of her ring finger. She winces in pain. Physiologically and neurologically, a lot happened—from the tissue damage caused by the metal splinter, to Freya's finally wincing from the sensation. But it only took milliseconds.

   Now suppose that while out foraging and hopping about, the wild rabbit mishops on the prickly side of a pinecone. The rabbit cries out a bit, winks hard, and hops off fast. A very similar physiological and neurological chain of events no doubt transpired from the mishop on the pinecone to hopping off fast in pain. But as interestingly similar as the wild rabbit's brain is to human Freya's, it is not plausible to think that both Freya and the wild rabbit entered into the same brain state. We do want to say they entered into the same *mental* state, however. That is, they were both in pain. Since the same pain state can be realized in multiple kinds of brains, we can say that mental states like pain are multiply realizable. This is bad news for the materialist; it looks like brain states and mental states come apart.

Avoiding Behaviourism

Now we look bleary-eyed in the direction of behaviourism. But here, too, we find a suspicious identity claim—this time between mental states, like Freya's belief that her house is gray, and behavioural states or dispositions to behave in certain circumstances. For

example, if Freya were asked what color her house is, she would be disposed to answer, "Gray." But just as with mental states and brain states, Freya's believing that her Colonial-period house is painted the original gray from when the house was first built and painted in 1810, and her dispositions to behave accordingly, come apart, showing that they could not be identical.

Suppose Freya wants to throw a housewarming party for herself and includes a colorful direction in the invitation that hers is the "only big gray Colonial on Jones St. Can't miss it." We say that Freya would not sincerely include such a thing if she did not believe it to be true. And we have no reason to suspect she is lying. We can go further. We want to say that it is her belief that her Colonial is big, is gray, and the only one like it on Jones Street that causes her, at least in part, to include that direction in the invitation. But if it is her mental state (her belief) that caused her behaviour, then the mental state and the behavioural state (her including the colorful direction in the invitation) cannot be strictly identical.

Freya might very well have been disposed to give just such a colorful direction to her home, given her beliefs, as the behaviourist would predict; and this disposition might even come with believing the things Freya does. But if we want to refer to Freya's beliefs in our explanation of her behaviour—and this is the sort of thing we do when we say our beliefs and other mental states cause our behaviour—then we must hold that they are distinct, since otherwise our causal explanation would be viciously circular.

It would be circular because the thing to be explained, her Colonial-describing behaviour, is the same thing as the thing that is supposed to causally explain it, her Colonial-descriptive beliefs; and the circle would be vicious because nothing would ever really get explained. So the behaviourist, like the materialist, seems to see an identity where there is none.

No Turning Back: The Mind is Natural

The goal is to formulate an alternative to the above two theories of mind that nevertheless both make a promise worth making: to treat the mind as something wholly a part of the natural world. From the failures of materialism and behaviourism, we must not turn back to a problematic Cartesian dualist view of mind and matter (discussed in Chapter 1), where it again would become utterly mysterious how Freya's beliefs about how her Colonial looks could possibly influence her physical behaviour, since her beliefs and physical behaviour exist on different planes of existence, as it were. But there is a third way to view beliefs like Freya's.

Functionalism as the Middle Path

Our way between the two monsters is to take seriously the perhaps dangerous idea that minds really are computing machines. In England, Alan Turing (1912-1954) laid the groundwork for such an idea with his monumental work on the nature of computing machines and intelligence (1936, 230-265; 1950, 433-460). Turing was able to conceive of a computing machine so powerful that it could successfully perform any computable function a human being could be said to carry out, whether consciously, as in the math classroom, or at the subconscious level, as in the many computations involved in navigating from one side to the other of one's room.

A Turing machine, as it came to be called, is an abstract computer model designed with the purpose of illustrating the limits of computability. Thinking creatures like human beings, of course, are not abstract things. Turing machines are not themselves thinking machines, but insofar as thinking states can be coherently understood as computational states, a Turing machine or Turing

machine-inspired model should provide an illuminating account of the mind.

Turing's ideas were developed in the United States by philosopher Hilary Putnam (1926-2016). Functionalism treats minds as natural phenomena contra Cartesian dualism; mental states, like pain, as multiply realizable, contra materialism; and mental states as causes of behaviour, contra behaviourism. In its simple form, it is the joint thesis that the mind is a functional system, kind of like an operating system of a computer, and mental states like beliefs, desires, and perceptual experiences are really just functional states, kind of like inputs and outputs in that operating system. Indeed, often this simple version of functionalism is known as "machine" or "input-output functionalism" to highlight just those mechanical features of the theory.

Nothing's Shocking: The Functionalist Mind is a Natural Mind

The functionalist says if we conceive of mental stuff in this way—namely, as fundamentally inputs and outputs in a complex, but wholly natural system—then we get to observe the reality of the mind, and the reality of our mental lives. We get to avoid any genuine worries about mental stuff being too spooky, or about how it could possibly interact with material stuff, as one might genuinely worry on a Cartesian dualist theory of mind, where we are asked to construe mental stuff and material stuff as fundamentally two kinds of substances. With functionalism, the how-possible question about interaction between the mental and material simply does not arise, no more than it would for the software and hardware interaction in computers, respectively. So, on the functionalist picture of the mind, the mysterious fog is lifted, and the way is clear.

Multiple Realizability

Let us use a thought experiment of our own to illustrate the functionalist's theory of mind. Imagine Freya cooks a warm Sunday breakfast for herself and sits on a patio table in the spring sun to enjoy it. Freya's belief that "my tofu scramble is on the table before me" is to be understood roughly like this: as the OUTPUT of one mental state, her seeing her breakfast on the table before her, and as the INPUT for others, including other beliefs Freya might have or come to have by deductive inference ("something is on the table before me," and so on and so forth) and behaviours (e.g., sticking a fork into that tofu scramble and scarfing it down). Note well: we have not mentioned anything here about the work Freya's sensory cortex or thalamus or the role the rods and cones in her retina are playing in getting her to believe what she does; her belief is identified only by its functional or causal role. This seems to imply that Freya's breakfast belief is multiply realizable, like pain is.

Recall our earlier discussion of the important difference between rabbit-brain stuff and human-brain stuff. Nevertheless, we wanted to say that both Freya and the wild rabbit could be in pain. We said pain, then, is multiply realizable. This is another way of saying that being in pain does not require any specific realization means, just some or other adequate means of realization. The point also strongly implies that the means of realization for Freya's breakfast belief, no less than her pain, need not be a brain state at all. This signals a major worry for the materialist. Since our beliefs, desires, and perceptual experiences are identified by their functional or causal role, the functionalist has no problem accounting for the multiple realizability of mental states.

Real Cause: The Functionalist Mind Causes Behaviour

Finally, we saw that our mental states cannot be counted as the causes of our behaviour on a behaviourist view, since on that view of mind, mental states are nothing over and above our behaviour (or, dispositions to behave in certain ways in certain circumstances). In an effort to disenchant the mind in general and individual minds in particular, and move mental states like beliefs and pain into scientific view, the behaviourist recoiled too far from spooky Cartesian dualism, leaving nothing in us to be the causes of our own behaviour. The functionalist understands, like the behaviourist, that there is a close connection between our beliefs, desires, and pains, on the one hand, and our behaviour, on the other. It is just that the connection is a functional, or causal, one, not one of identity. Since mental states (like Freya's belief that "my tofu scramble is on the table before me") are identified with their functional or causal role in the larger functional system of inputs and outputs, other mental states and behavioural states, the functionalist has no problem accounting for mental states playing a causal role in the explanations we give of our own behaviour. On the functionalist theory of mind, mental states are real causes of behaviour.

Objections to Functionalism

Now that we have seen some of the major points in favor of the theory, let us have a look at some of the worries that have been raised against functionalism.

The Chinese Room

John Searle argues against a version of functionalism he calls "strong" artificial intelligence, or "strong AI" In "Minds, Brains and Programs," Searle develops a thought experiment designed to show that having the right inputs and outputs is not sufficient for having mental states, as the functionalist claims (1980). The specific issue concerns what is required to understand Chinese.

Imagine someone who does not understand Chinese is put in a room and tasked with sorting Chinese symbols in response to other Chinese symbols, according to purely formal rules given in an English-language manual. So, for example, one person can write some Chinese symbols on a card, place it in a basket on a conveyor belt which leads into and out from the little room you are in. Once you receive it, you look at the shape of the symbol, find it in the manual, and read which Chinese symbols to find in the other basket to send back out. Imagine further that you get very good at this manipulation of symbols, so good in fact that you can fool fluent Chinese speakers with the responses you give. To them, you function every bit like you understand Chinese. It appears, however, you have no true understanding at all. Therefore, Searle concludes, functioning in the right way is not sufficient for having mental states.

The functionalist has replied that, of course, as the thought experiment is described, the person in the room does not understand Chinese. But also as the case is described, the person in the room is just a piece of the whole functional system. Indeed, it is the system that functions to understand Chinese, not just one part. So it is the whole system, in this case, the whole room, including the person manipulating the symbols and the instruction manual (the "program"), that understands Chinese.

The Problem of Qualia

The splinter Freya picked up from her D string caused her a bit of pain, and perhaps more so for the behaviourist, as we saw earlier. One major worry for the functionalist is that there seems to be more to Freya's pain than its just being the putative cause of some pain-related behaviour, where this cause is understood to be another mental state, presumably, not identified with pain at all. (Remember, the functionalist wishes to avoid the vicious circularity that plagued the behaviourist's explanations of behaviour.)

There is an undeniable sensation to pain: it is something you feel. In fact, some might argue that at the conscious level, that is all there is to pain. Sure, there is the detection of tissue damage and the host physiological and neurological events transpiring, and yes, there is the pain-related behaviour, too. However, we must not leave out of our explanation of pain the feel of pain. Philosophers call the feeling aspect of some mental states like pain fundamentally qualitative states. Other qualitative mental states might include experiences of colored objects, such as those a person with normal color vision has every day.

In seeing a Granny Smith apple in the basket on a dining room table, she has a visual experience as of a green object. But the functionalist can only talk about the experience in terms of the function or causal role it plays. So, for example, the functionalist can speak to Freya's green experience as being the cause of her belief that she sees a green apple in the basket. But the functionalist cannot speak to the feeling Freya (or any of us) has in seeing a ripe green Granny Smith. We think there is a corresponding feeling to color experiences like Freya's over and above whatever beliefs they might go on to cause us to have. Since mental states like pain and color experiences are identified solely by their functional role, the functionalist seems without the resources to account for these qualitative mental states.

The functionalist might reply by offering a treatment of qualia in

terms of what such aspects of experience function to do for us. The vivid, ripe greenness of the Granny Smith functions to inform Freya about a source of food in a way that pulls her visual attention to it. Freya's color experiences allow her to form accurate beliefs about the objects in her immediate environment. It is certainly true that ordinary visual experience provide us with beautiful moments in our lives. However, they likely function to do much more besides. Likewise, it is more likely that there is a function for the qualitative or feeling aspects of some mental states, and that these aspects can be understood in terms of their functions, than it is that these aspects are free-floating above the causal order of things. So, the functionalist who wishes to try to account for qualia need not remain silent on the issue.

Conclusion

We have not considered all the possible objections to functionalism, nor have we considered more sophisticated versions of functionalism that aim to get around the more pernicious objections we have considered. The idea that minds really are kinds of computing machines is still very much alive and as controversial as ever. Taking that idea seriously means having to wrestle with a host of questions at the intersection of philosophy of mind, philosophy of action, and personal identity.

In what sense is Freya truly an *agent* of her own actions, if we merely cite a cold input to explain some behaviour of hers? That is to say, how does Freya *avow* her own beliefs on a merely functionalist view? If minds are kinds of computers, then what does that make thinking creatures like Freya? Kinds of robots, albeit sophisticated ones? These and other difficult questions will need to be answered satisfactorily before many philosophers will be content with a functionalist theory of mind. For other philosophers, a start down the right path, away from Cartesian dualism and between

the two terrors of materialism and behaviourism, has already been made.

References

Putnam, Hilary. (1960) 1975. "Minds and Machines." Reprinted in *Mind, Language, and Reality*, 362-385. Cambridge: Cambridge University Press.

Searle, John. 1980. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3(3): 417-457.

Turing, Alan, M. 1936. "On Computable Numbers, with an Application to the Entscheidungsproblem." *Proceedings of the London Mathematical Society* 42 (1): 230-265.

Turing, Alan, M. 1950. "Computing Machinery and Intelligence." *Mind* 49: 433-460.

Further Reading

Block, Ned. 1980a. *Readings in the Philosophy of Psychology, Volumes 1 and 2*. Cambridge, MA: Harvard University Press.

Block, Ned. 1980b. "Troubles With Functionalism." In Block 1980a, 268-305.

Gendler, Tamar. 2008. "Belief and Alief." *Journal of Philosophy* 105(10): 634-663.

Jackson, Frank. 1982. "Epiphenomenal Qualia." *Philosophical Quarterly* 32: 127-136.

Lewis, David. 1972. "Psychophysical and Theoretical Identifications." In Block 1980a, 207-215.

Lewis, David. 1980. "Mad Pain and Martian Pain." In Block 1980, 216-222.

Nagel, Thomas. 1974. "What Is It Like To Be a Bat?" *Philosophical Review* 83: 435-450.

Putnam, Hilary. 1963. "Brains and Behavior." Reprinted in Putnam 1975b, 325-341.

Putnam, Hilary. 1967. "The Nature of Mental States." Reprinted in Putnam 1975b, 429-440.

Putnam, Hilary. 1973. "Philosophy and our Mental Life." Reprinted in Putnam 1975b, 291-303.

Putnam, Hilary. 1975a. "The Meaning of 'Meaning.'" Reprinted in Putnam 1975b, 215-271.

Putnam, Hilary. 1975b. *Mind, Language, and Reality*. Cambridge: Cambridge University Press.

Shoemaker, Sydney. 1984. *Identity, Cause, and Mind*. Cambridge: Cambridge University Press.

Shoemaker, Sydney. 1996. *The First-Person Perspective and Other Essays*. Cambridge: Cambridge University Press.

27. Arguments For and Against Conscious Computers

No Turning Back: The Mind is Natural

The goal is to formulate an alternative to the above two theories of mind that nevertheless both make a promise worth making: to treat the mind as something wholly a part of the natural world. From the failures of materialism and behaviorism, we must not turn back to a problematic Cartesian dualist view of mind and matter (discussed in Chapter 1), where it again would become utterly mysterious how Freya's beliefs about how her Colonial looks could possibly influence her physical behavior, since her beliefs and physical behavior exist on different planes of existence, as it were. But there is a third way to view beliefs like Freya's.

Functionalism as the Middle Path

Our way between the two monsters is to take seriously the perhaps dangerous idea that minds really are computing machines. In England, Alan Turing (1912-1954) laid the groundwork for such an idea with his monumental work on the nature of computing machines and intelligence (1936, 230-265; 1950, 433-460). Turing was able to conceive of a computing machine so powerful that it could successfully perform any computable function a human being could be said to carry out, whether consciously, as in the math classroom, or at the subconscious level, as in the many

computations involved in navigating from one side to the other of one's room.

A Turing machine, as it came to be called, is an abstract computer model designed with the purpose of illustrating the limits of computability. Thinking creatures like human beings, of course, are not abstract things. Turing machines are not themselves thinking machines, but insofar as thinking states can be coherently understood as computational states, a Turing machine or Turing machine-inspired model should provide an illuminating account of the mind.

Turing's ideas were developed in the United States by philosopher Hilary Putnam (1926-2016). Functionalism treats minds as natural phenomena contra Cartesian dualism; mental states, like pain, as multiply realizable, contra materialism; and mental states as causes of behavior, contra behaviorism. In its simple form, it is the joint thesis that the mind is a functional system, kind of like an operating system of a computer, and mental states like beliefs, desires, and perceptual experiences are really just functional states, kind of like inputs and outputs in that operating system. Indeed, often this simple version of functionalism is known as "machine" or "input-output functionalism" to highlight just those mechanical features of the theory.

Nothing's Shocking: The Functionalist Mind is a Natural Mind

The functionalist says if we conceive of mental stuff in this way—namely, as fundamentally inputs and outputs in a complex, but wholly natural system—then we get to observe the reality of the mind, and the reality of our mental lives. We get to avoid any genuine worries about mental stuff being too spooky, or about how it could possibly interact with material stuff, as one might genuinely worry on a Cartesian dualist theory of mind, where we are asked to

construe mental stuff and material stuff as fundamentally two kinds of substances. With functionalism, the how-possible question about interaction between the mental and material simply does not arise, no more than it would for the software and hardware interaction in computers, respectively. So, on the functionalist picture of the mind, the mysterious fog is lifted, and the way is clear.

28. Conscious Computers?

The functionalist might reply by offering a treatment of qualia in terms of what such aspects of experience function to do for us. The vivid, ripe greenness of the Granny Smith functions to inform Freya about a source of food in a way that pulls her visual attention to it. Freya's color experiences allow her to form accurate beliefs about the objects in her immediate environment. It is certainly true that ordinary visual experience provide us with beautiful moments in our lives. However, they likely function to do much more besides. Likewise, it is more likely that there is a function for the qualitative or feeling aspects of some mental states, and that these aspects can be understood in terms of their functions, than it is that these aspects are free-floating above the causal order of things. So, the functionalist who wishes to try to account for qualia need not remain silent on the issue.

Conclusion

We have not considered all the possible objections to functionalism, nor have we considered more sophisticated versions of functionalism that aim to get around the more pernicious objections we have considered. The idea that minds really are kinds of computing machines is still very much alive and as controversial as ever. Taking that idea seriously means having to wrestle with a host of questions at the intersection of philosophy of mind, philosophy of action, and personal identity.

In what sense is Freya truly an *agent* of her own actions, if we merely cite a cold input to explain some behavior of hers? That is to say, how does Freya *avow* her own beliefs on a merely functionalist view? If minds are kinds of computers, then what does that make

thinking creatures like Freya? Kinds of robots, albeit sophisticated ones? These and other difficult questions will need to be answered satisfactorily before many philosophers will be content with a functionalist theory of mind. For other philosophers, a start down the right path, away from Cartesian dualism and between the two terrors of materialism and behaviorism, has already been made.

References

Putnam, Hilary. (1960) 1975. "Minds and Machines." Reprinted in *Mind, Language, and Reality*, 362-385. Cambridge: Cambridge University Press.

Searle, John. 1980. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3(3): 417-457.

Turing, Alan, M. 1936. "On Computable Numbers, with an Application to the Entscheidungsproblem." *Proceedings of the London Mathematical Society* 42 (1): 230-265.

Turing, Alan, M. 1950. "Computing Machinery and Intelligence." *Mind* 49: 433-460.

Further Reading

Block, Ned. 1980a. *Readings in the Philosophy of Psychology, Volumes 1 and 2*. Cambridge, MA: Harvard University Press.

Block, Ned. 1980b. "Troubles With Functionalism." In Block 1980a, 268-305.

Gendler, Tamar. 2008. "Belief and Alief." *Journal of Philosophy* 105(10): 634-663.

Jackson, Frank. 1982. "Epiphenomenal Qualia." *Philosophical Quarterly* 32: 127-136.

Lewis, David. 1972. "Psychophysical and Theoretical Identifications." In Block 1980a, 207-215.

Lewis, David. 1980. "Mad Pain and Martian Pain." In Block 1980, 216-222.

Nagel, Thomas. 1974. "What Is It Like To Be a Bat?" *Philosophical Review* 83: 435-450.

Putnam, Hilary. 1963. "Brains and Behavior." Reprinted in Putnam 1975b, 325-341.

Putnam, Hilary. 1967. "The Nature of Mental States." Reprinted in Putnam 1975b, 429-440.

Putnam, Hilary. 1973. "Philosophy and our Mental Life." Reprinted in Putnam 1975b, 291-303.

Putnam, Hilary. 1975a. "The Meaning of 'Meaning.'" Reprinted in Putnam 1975b, 215-271.

Putnam, Hilary. 1975b. *Mind, Language, and Reality*. Cambridge: Cambridge University Press.

Shoemaker, Sydney. 1984. *Identity, Cause, and Mind*. Cambridge: Cambridge University Press.

Shoemaker, Sydney. 1996. *The First-Person Perspective and Other Essays*. Cambridge: Cambridge University Press.

29. Attribution

The material in this chapter is adapted from previously published content:

Blum, P.R. (2019) Substance Dualism in Descartes. In Brandon, B. & Hendricks, C., eds. *Introduction to Philosophy: Philosophy of the Mind* (pp. 1.5). Rebus Community. https://press.rebus.community/intro-to-phil-of-mind/ CC BY.

Newman, J. (2019) Functionalism. In Brandon, B. & Hendricks, C., eds. *Introduction to Philosophy: Philosophy of the Mind* (pp. 3.22-28). Rebus Community. https://press.rebus.community/intro-to-phil-of-mind/ CC BY.

Shevlin, H. (2019) Qualia and Raw Feels. In Brandon, B. & Hendricks, C., eds. *Introduction to Philosophy: Philosophy of the Mind* (pp. 5.38). Rebus Community. https://press.rebus.community/intro-to-phil-of-mind/ CC BY.

Vintiadis, E. (2019) Property Dualism. In Brandon, B. & Hendricks, C., eds. *Introduction to Philosophy: Philosophy of the Mind* (pp. 4.3). Rebus Community. https://press.rebus.community/intro-to-phil-of-mind/ CC BY.

FREE WILL

30. Existence of Free Will

Determinism and Freedom

Determinism and free will are often thought to be in deep conflict. Whether or not this is true has a lot to do with what is meant by determinism and an account of what free will requires.

First of all, determinism is not the view that free actions are impossible. Rather, determinism is the view that at any one time, only one future is physically possible. To be a little more specific, determinism is the view that a complete description of the past along with a complete account of the relevant laws of nature logically entails all future events.[1]

Indeterminism is simply the denial of determinism. If determinism is incompatible with free will, it will be because free actions are only possible in worlds in which more than one future is physically possible at any one moment in time. While it might be true that free will requires indeterminism, it's not true merely by definition. A further argument is needed and this suggests that it is at least possible that people could sometimes exercise the control necessary for morally responsible action, even if we live in a deterministic world.

It is worth saying something about fatalism before we move on. It is really easy to mistake determinism for fatalism, and fatalism does seem to be in straightforward conflict with free will. Fatalism is the view that we are powerless to do anything other than what we actually do. If fatalism is true, then nothing that we try or think or intend or believe or decide has any causal effect or relevance as to what we actually end up doing.

But note that determinism need not entail fatalism. Determinism is a claim about what is logically entailed by the rules/laws governing a world and the past of said world. It is not the claim that

we lack the power to do other than what we actually were already going to do. Nor is it the view that we fail to be an important part of the causal story for why we do what we do. And this distinction may allow some room for freedom, even in deterministic worlds.

An example will be helpful here. We know that the boiling point for water is 100°C. Suppose we know in both a deterministic world and a fatalistic world that my pot of water will be boiling at 11:22am today. Determinism makes the claim that if I take a pot of water and I put it on my stove, and heat it to 100°C, it will boil. This is because the laws of nature (in this case, water that is heated to 100°C will boil) and the events of the past (I put a pot of water on a hot stove) bring about the boiling water. But fatalism makes a different claim. If my pot of water is fated to boil at 11:22am today, then no matter what I or anyone does, my pot of water will boil at exactly 11:22am today. I could try to empty the pot of water out at 11:21. I could try to take the pot as far away from a heating source as possible. Nonetheless, my pot of water will be boiling at 11:22 precisely because it was fated that this would happen. Under fatalism, the future is fixed or preordained, but this need not be the case in a deterministic world. Under determinism, the future is a certain way because of the past and the rules governing said world. If we know that a pot of water will boil at 11:22am in a deterministic world, it's because we know that the various causal conditions will hold in our world such that at 11:22 my pot of water will have been put on a heat source and brought to 100°C. Our deliberations, our choices, and our free actions may very well be part of the process that brings a pot of water to the boiling point in a deterministic world, whereas these are clearly irrelevant in fatalistic ones.

Three Views of Freedom

Most accounts of freedom fall into one of three camps. Some people take freedom to require merely the ability to "do what you want to

do." For example, if you wanted to walk across the room, right now, and you also had the ability, right now, to walk across the room, you would be free as you could do exactly what you want to do. We will call this easy freedom.

Others view freedom on the infamous "Garden of Forking Paths" model. For these people, free action requires more than merely the ability to do what you want to do. It also requires that you have the ability to do otherwise than what you actually did. So, If Anya is free when she decides to take a sip from her coffee, on this view, it must be the case that Anya could have refrained from sipping her coffee. The key to freedom, then, is alternative possibilities and we will call this the alternative possibilities view of free action.

Finally, some people envision freedom as requiring, not alternative possibilities but the right kind of relationship between the antecedent sources of our actions and the actions that we actually perform. Sometimes this view is explained by saying that the free agent is the source, perhaps even the ultimate source of her action. We will call this kind of view a source view of freedom.

Now, the key question we want to focus on is whether or not any of these three models of freedom are compatible with determinism. It could turn out that all three kinds of freedom are ruled out by determinism, so that the only way freedom is possible is if determinism is false. If you believe that determinism rules out free action, you endorse a view called incompatibilism. But it could turn out that one or all three of these models of freedom are compatible with determinism. If you believe that free action is compatible with determinism, you are a compatibilist.

Let us consider compatibilist views of freedom and two of the most formidable challenges that compatibilists face: the consequence argument and the ultimacy argument.

Begin with easy freedom. Is easy freedom compatible with determinism? A group of philosophers called classic compatibilists certainly thought so.[2] They argued that free will requires merely the ability for an agent to act without external hindrance. Suppose, right now, you want to put your textbook down and grab a cup of

coffee. Even if determinism is true, you probably, right now, can do exactly that. You can put your textbook down, walk to the nearest Starbucks, and buy an overpriced cup of coffee. Nothing is stopping you from doing what you want to do. Determinism does not seem to be posing any threat to your ability to do what you want to do right now. If you want to stop reading and grab a coffee, you can. But, by contrast, if someone had chained you to the chair you are sitting in, things would be a bit different. Even if you wanted to grab a cup of coffee, you would not be able to. You would lack the ability to do so. You would not be free to do what you want to do. This has nothing to do with determinism, of course. It is not the fact that you might be living in a deterministic world that is threatening your free will. It is that an external hindrance (the chains holding you to your chair) is stopping from you doing what you want to do. So, if what we mean by freedom is easy freedom, it looks like freedom really is compatible with determinism.

Easy freedom has run into some rather compelling opposition, and most philosophers today agree that a plausible account of easy freedom is not likely. But, by far, the most compelling challenge the view faces can be seen in the consequence argument.[3] The consequence argument is as follows:

1. If determinism is true, then all human actions are consequences of past events and the laws of nature.
2. No human can do other than they actually do except by changing the laws of nature or changing the past.
3. No human can change the laws of nature or the past.

> 4. If determinism is true, no human has free will.

This is a powerful argument. It is very difficult to see where this argument goes wrong, if it goes wrong. The first premise is merely a restatement of determinism. The second premise ties the ability to do otherwise to the ability to change the past or the laws of nature, and the third premise points out the very reasonable assumption that humans are unable to modify the laws of nature or the past.

This argument effectively devastates easy freedom by proposing that we never act without external hindrances precisely because our actions are caused by past events and the laws of nature in such a way that we not able to contribute anything to the causal production of our actions. This argument also seems to pose a deeper problem for freedom in deterministic worlds. If this argument works, it establishes that, given determinism, we are powerless to do otherwise, and to the extent that freedom requires the ability to do otherwise, this argument seems to rule out free action. Note that if this argument works, it poses a challenge for both the easy and alternative possibilities view of free will.

How might someone respond to this argument? First, suppose you adopt an alternative possibilities view of freedom and believe that the ability to do otherwise is what is needed for genuine free will. What you would need to show is that alternative possibilities, properly understood, are not incompatible with determinism. Perhaps you might argue that if we understand the ability to do otherwise properly we will see that we actually do have the ability to change the laws of nature or the past.

That might sound counterintuitive. How could it possibly be the case that a mere mortal could change the laws of nature or the past? Think back to Quinn's decision to spend the night before her exam out with friends instead of studying. When she shows up to her

exam exhausted, and she starts blaming herself, she might say, "Why did I go out? That was dumb! I could have stayed home and studied." And she is sort of right that she could have stayed home. She had the general ability to stay home and study. It is just that if she had stayed home and studied the past would be slightly different or the laws of nature would be slightly different. What this points to is that there might be a way of cashing out the ability to do otherwise that is compatible with determinism and does allow for an agent to kind of change the past or even the laws of nature.[4]

But suppose we grant that the consequence argument demonstrates that determinism really does rule out alternative possibilities. Does that mean we must abandon the alternative possibilities view of freedom? Well, not necessarily. You could instead argue that free will is possible, provided determinism is false.[5] That is a big if, of course, but maybe determinism will turn out to be false.

What if determinism turns out to be true? Should we give up, then, and concede that there is no free will? Well, that might be too quick. A second response to the consequence argument is available. All you need to do is deny that freedom requires the ability to do otherwise.

In 1969, Harry Frankfurt proposed an influential thought experiment that demonstrated that free will might not require alternative possibilities at all (Frankfurt [1969] 1988). If he's right about this, then the consequence argument, while compelling, does not demonstrate that no one lacks free will in deterministic worlds, because free will does not require the ability to do otherwise. It merely requires that agents be the source of their actions in the right kind of way. But we're getting ahead of ourselves. Here is a simplified paraphrase of Frankfurt's case:

> Black wants Jones to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid unnecessary work. So he waits until Jones is about to make up his mind what to do, and he

does nothing unless it is clear to him (Black is an excellent judge of such things) that Jones is going to decide not to do what Black wants him to do. If it does become clear that Jones is going to decide to do something other than what Black wanted him to do, Black will intervene, and ensure that Jones decides to do, and does do, exactly what Black wanted him to do. Whatever Jones' initial preferences and inclinations, then, Black will have his way. As it turns out, Jones decides, on his own, to do the action that Black wanted him to perform. So, even though Black was entirely prepared to intervene, and could have intervened, to guarantee that Jones would perform the action, Black never actually has to intervene because Jones decided, for reasons of his own, to perform the exact action that Black wanted him to perform. (Frankfurt [1969] 1988, 6-7)

Now, what is going on here? Jones is overdetermined to perform a specific act. No matter what happens, no matter what Jones initially decides or wants to do, he is going to perform the action Black wants him to perform. He absolutely cannot do otherwise. But note that there seems to be a crucial difference between the case in which Jones decides on his own and for his own reasons to perform the action Black wanted him to perform and the case in which Jones would have refrained from performing the action were it not for Black intervening to force him to perform the action. In the first case, Jones is the source of his action. It the thing he decided to do and he does it for his own reasons. But in the second case, Jones is not the source of his actions. Black is. This distinction, thought Frankfurt, should be at the heart of discussions of free will and moral responsibility. The control required for moral responsibility is not the ability to do otherwise (Frankfurt [1969] 1988, 9-10).

If alternative possibilities are not what free will requires, what kind of control is needed for free action? Here we have the third view of freedom we started with: free will as the ability to be the source of your actions in the right kind of way. Source compatibilists

argue that this ability is not threatened by determinism, and building off of Frankfurt's insight, have gone on to develop nuanced, often radically divergent source accounts of freedom.[6] Should we conclude, then, that provided freedom does not require alternative possibilities that it is compatible with determinism?[7] Again, that would be too quick. Source compatibilists have reason to be particularly worried about an argument developed by Galen Strawson called the ultimacy argument (Strawson [1994] 2003, 212-228).

Rather than trying to establish that determinism rules out alternative possibilities, Strawson tried to show that determinism rules out the possibility of being the ultimate source of your actions. While this is a problem for anyone who tries to establish that free will is compatible with determinism, it is particularly worrying for source compatibilists as they've tied freedom to an agent's ability to be source of its actions. Here is the argument:

1. A person acts of her own free will only if she is the act's ultimate source.
2. If determinism is true, no one is the ultimate source of her actions.

---

3. Therefore, if determinism is true, no one acts of her own free will. (McKenna and Pereboom 2016, 148)[8]

This argument requires some unpacking. First of all, Strawson argues that for any given situation, we do what we do because of the way we are ([1994] 2003, 219). When Quinn decides to go out

with her friends rather than study, she does so because of the way she is. She prioritizes a night with her friends over studying, at least on that fateful night before her exam. If Quinn had stayed in and studied, it would be because she was slightly different, at least that night. She would be such that she prioritized studying for her exam over a night out. But this applies to any decision we make in our lives. We decide to do what we do because of how we already are.

But if what we do is because of the way we are, then in order to be responsible for our actions, we need to be the source of how we are, at least in the relevant mental respects (Strawson [1994] 2003, 219). There is the first premise. But here comes the rub: the way we are is a product of factors beyond our control such as the past and the laws of nature ([1994] 2003, 219; 222-223). The fact that Quinn is such that she prioritizes a night with friends over studying is due to her past and the relevant laws of nature. It is not up to her that she is the way she is. It is ultimately factors extending well beyond her, possibly all the way back to the initial conditions of the universe that account for why she is the way she is that night. And to the extent that this is compelling, the ultimate source of Quinn's decision to go out is not her. Rather, it is some condition of the universe external to her. And therefore, Quinn is not free.

Once again, this is a difficult argument to respond to. You might note that "ultimate source" is ambiguous and needing further clarification. Some compatibilists have pointed this out and argued that once we start developing careful accounts of what it means to be the source of our actions, we will see that the relevant notion of source-hood is compatible with determinism.

For example, while it may be true that no one is the ultimate cause of their actions in deterministic worlds precisely because the ultimate source of all actions will extend back to the initial conditions of the universe, we can still be a mediated source of our actions in the sense required for moral responsibility. Provided the actual source of our action involves a sophisticated enough set of capacities for it to make sense to view us as the source of our actions, we could still be the source of our actions, in the

relevant sense (McKenna and Pereboom 2016, 154). After all, even if determinism is true, we still act for reasons. We still contemplate what to do and weigh reasons for and against various actions, and we still are concerned with whether or not the actions we are considering reflect our desires, our goals, our projects, and our plans. And you might think that if our actions stem from a history that includes us bringing all the features of our agency to bear upon the decision that is the proximal cause of our action, that this causal history is one in which we are the source of our actions in the way that is really relevant to identifying whether or not we are acting freely.

Others have noted that even if it is true that Quinn is not directly free in regard to the beliefs and desires that suggest she should go out with her friends rather than study (they are the product of factors beyond her control such as her upbringing, her environment, her genetics, or maybe even random luck), this need not imply that she lacks control as to whether or not she chooses to act upon them.[9] Perhaps it is the case that even though how we are may be due to factors beyond our control, nonetheless, we are still the source of what we do because it is still, even under determinism, up to us as to whether we choose to exercise control over our conduct.

Free Will and the Sciences

Many challenges to free will come, not from philosophy, but from the sciences. There are two main scientific arguments against free will, one coming from neuroscience and one coming from the social sciences. The concern coming from research in the neurosciences is that some empirical results suggest that all our choices are the result of unconscious brain processes, and to the extent choices must be consciously made to be free choices, it seems that we never make a conscious free choice.

The classic studies motivating a picture of human action in which unconscious brain processes are doing the bulk of the causal work for action were conducted by Benjamin Libet. Libet's experiments involved subjects being asked to flex their wrists whenever they felt the urge to do so. Subjects were asked to note the location of a clock hand on a modified clock when they became aware of the urge to act. While doing this their brain activity was being scanned using EEG technology. What Libet noted is that around 550 milliseconds before a subject acted, a readiness potential (increased brain activity) would be measured by the EEG technology. But subjects were reporting awareness of an urge to flex their wrist around 200 milliseconds before they acted (Libet 1985).

This painted a strange picture of human action. If conscious intentions were the cause of our actions, you may expect to see a causal story in which the conscious awareness of an urge to flex your wrist shows up first, then a ramping up of brain activity, and finally an action. But Libet's studies showed a causal story in which an action starts with unconscious brain activity, the subject later becomes consciously aware that they are about to act, and then the action happens. The conscious awareness of action seemed to be a byproduct of the actual unconscious process that was causing the action. It was not the cause of the action itself. And this result suggests that unconscious brain processes, not conscious ones, are the real causes of our actions. To the extent that free action requires our conscious decisions to be the initiating causes of our actions, it looks like we may never act freely.

While this research is intriguing, it probably does not establish that we are not free. Alfred Mele is a philosopher who has been heavily critical of these studies. He raises three main objections to the conclusions drawn from these arguments.

First, Mele points out that self-reports are notoriously unreliable (2009, 60-64). Conscious perception takes time, and we are talking about milliseconds. The actual location of the clock hand is probably much closer to 550 milliseconds when the agent "intends" or has

the "urge" to act than it is to 200 milliseconds. So, there's some concerns about experimental design here.

Second, an assumption behind these experiments is that what is going on at 550 milliseconds is that a decision is being made to flex the wrist (Mele 2014, 11). We might challenge this assumption. Libet ran some variants of his experiment in which he asked subjects to prepare to flex their wrist but to stop themselves from doing so. So, basically, subjects simply sat there in the chair and did nothing. Libet interpreted the results of these experiments as showing that we might not have a free will, but we certainly have a "free won't" because we seem capable of consciously vetoing or stopping an action, even if that action might be initiated by unconscious processes (2014, 12-13). Mele points out that what might be going on in these scenarios is that the real intention to act or not act is what happens consciously at 200 milliseconds, and if so, there is little reason to think these experiments are demonstrating that we lack free will (2014, 13).

Finally, Mele notes that while it may be the case that some of our decisions and actions look like the wrist-flicking actions Libet was studying, it is doubtful that all or even most of our decisions are like this (2014, 15). When we think about free will, we rarely think of actions like wrist-flicking. Free actions are typically much more complex and they are often the kind of thing where the decision to do something extends across time. For example, your decision about what to major in at college or even where to study was probably made over a period of months, even years. And that decision probably involved periods of both conscious and unconscious cognition. Why think that a free choice cannot involve some components that are unconscious?

A separate line of attack on free will comes from the situationist literature in the social sciences (particularly social psychology). There is a growing body of research suggesting that situational and environmental factors profoundly influence human behavior, perhaps in ways that undermine free will (Mele 2014, 72).

Many of the experiments in the situationist literature are among

the most vivid and disturbing in all of social psychology. Stanley Milgram, for example, conducted a series of experiments on obedience in which ordinary people were asked to administer potentially lethal voltages of electricity to an innocent subject in order to advance scientific research, and the vast majority of people did so![10] And in Milgram's experiments, what affected whether or not subjects were willing to administer the shocks were minor, seemingly insignificant environmental factors such as whether the person running the experiment looked professional or not (Milgram 1963).

What experiments like Milgram's obedience experiments might show is that it is our situations, our environments that are the real causes of our actions, not our conscious, reflective choices. And this may pose a threat to free will. Should we take this kind of research as threatening freedom?

Many philosophers would resist concluding that free will does not exist on the basis of these kinds of experiments. Typically, not everyone who takes part in situationist studies is unable to resist the situational influences they are subject to. And it appears to be the case that when we are aware of situational influences, we are more likely to resist them. Perhaps the right way to think about this research is that there all sorts of situations that can influence us in ways that we may not consciously endorse, but that nonetheless, we are still capable of avoiding these effects when we are actively trying to do so. For example, the brain sciences have made many of us vividly aware of a whole host of cognitive biases and situational influences that humans are typically subject to and yet, when we are aware of these influences, we are less susceptible to them. The more modest conclusion to draw here is not that we lack free will, but that exercising control over our actions is much more difficult than many of us believe it to be. We are certainly influenced by the world we are a part of, but to be influenced by the world is different from being determined by it, and this may allow us to, at least sometimes, exercise some control over the actions we perform.

No one knows yet whether or not humans sometimes exercise the

control over their actions required for moral responsibility. And so I leave it to you, dear reader: Are you free?

Chapter Notes

1. I have hidden some complexity here. I have defined determinism in terms of logical entailment. Sometimes people talk about determinism as a causal relationship. For our purposes, this distinction is not relevant, and if it is easier for you to make sense of determinism by thinking of the past and the laws of nature causing all future events, that is perfectly acceptable to do.

2. Two of the more well-known classic compatibilists include Thomas Hobbes and David Hume. See: Hobbes, Thomas, (1651) 1994, *Leviathan*, ed. Edwin Curley, Canada: Hackett Publishing Company; and Hume, David, (1739) 1978, *A Treatise of Human Nature*, Oxford: Oxford University Press.

3. For an earlier version of this argument see: Ginet, Carl, 1966, "Might We Have No Choice?" in *Freedom and Determinism*, ed. Keith Lehrer, 87-104, Random House.

4. For two notable attempts to respond to the consequence argument by claiming that humans can change the past or the laws of nature see: Fischer, John Martin, 1994, *The Metaphysics of Free Will*, Oxford: Blackwell Publishers; and Lewis, David, 1981, "Are We Free to Break the Laws?" *Theoria* 47: 113-21.

5. Many philosophers try to develop views of freedom on the assumption that determinism is incompatible with free action. The view that freedom is possible, provided determinism is false is called Libertarianism. For more on Libertarian views of freedom, see: Clarke, Randolph and Justin Capes, 2017, "Incompatibilist (Nondeterministic) Theories of Free Will," *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/incompatibilism-

theories/.

6. For elaboration on recent compatibilist views of freedom, see McKenna, Michael and D. Justin Coates, 2015, "Compatibilism," *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/compatibilism/.

7. You might be unimpressed by the way source compatibilists understand the ability to be the source of your actions. For example, you might that what it means to be the source of your actions is to be the ultimate cause of your actions. Or maybe you think that to genuinely be the source of your actions you need to be the agent-cause of your actions. Those are both reasonable positions to adopt. Typically, people who understand free will as requiring either of these abilities believe that free will is incompatible with determinism. That said, there are many Libertarian views of free will that try to develop a plausible account of agent causation. These views are called Agent-Causal Libertarianism. See: Clarke, Randolph and Justin Capes, 2017, "Incompatibilist (Nondeterministic) Theories of Free Will," *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/incompatibilism-theories/.

8. As with most philosophical arguments, the ultimacy argument has been formulated in a number of different ways. In Galen Strawson's original paper he gives three different versions of the argument, one of which has eight premises and one that has ten premises. A full treatment of either of those versions of this argument would require more time and space than we have available here. I have chosen to use the McKenna/Pereboom formulation of the argument due its simplicity and their clear presentation of the central issues raised by the argument.

9. For two attempts to respond to the ultimacy argument in this way, see: Mele, Alfred, 1995, *Autonomous Agents*, New York: Oxford University Press; and McKenna, Michael, 2008, "Ultimacy & Sweet Jane" in Nick Trakakis and Daniel Cohen,

eds, *Essays on Free Will and Moral Responsibility*, Newcastle: Cambridge Scholars Publishing: 186-208.

10. Fortunately, no real shocks were administered. The subjects merely believed they were doing so.

References

Frankfurt, Harry. (1969) 1988. "Alternative Possibilities and moral responsibility." In *The Importance of What We Care About: Philosophical Essays*, 10th ed. New York: Cambridge University Press.

Libet, Benjamin. 1985. "Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action." *Behavioral and Brain Sciences* 8: 529-566.

McKenna, Michael and Derk Pereboom. 2016. *Free Will: A Contemporary Introduction.* New York: Routledge.

Mele, Alfred. 2014. *Free: Why Science Hasn't Disproved Free Will.* Oxford: Oxford University Press.

Mele, Alfred. 2009. *Effective Intentions: The Power of Conscious Will.* Oxford: Oxford University Press.

Milgram, Stanley. 1963. "Behavioral Study of Obedience." *The Journal of Abnormal and Social Psychology* 67: 371-378.

Strawson, Galen. (1994) 2003. "The Impossibility of Moral Responsibility." In *Free Will*, 2nd ed. Edited by Gary Watson, 212-228. Oxford: Oxford University Press.

van Inwagen, Peter. 1983. *An Essay on Free Will.* Oxford: Clarendon Press.

Further Reading

Deery, Oisin and Paul Russell, eds. 2013. *The Philosophy of Free Will:*

Essential Readings from the Contemporary Debates. New York: Oxford University Press.

Mele, Alfred. 2006. *Free Will and Luck.* New York: Oxford University Press.

Attribution

This section is composed of text taken from *Chapter 8 Freedom of the Will* created by Daniel Haas in *Introduction to Philosophy: Philosophy of Mind*, edited by Heather Salazar and Christina Hendricks, and produced with support from the Rebus Community. The original is freely available under the terms of the CC BY 4.0 license at https://press.rebus.community/intro-to-phil-of-mind/. The material is presented in its original form, with the exception of the removal of introductory material *Introduction: Are We Free?*

31. Argument Against Free Will

   This argument requires some unpacking. First of all, Strawson argues that for any given situation, we do what we do because of the way we are ([1994] 2003, 219). When Quinn decides to go out with her friends rather than study, she does so because of the way she is. She prioritizes a night with her friends over studying, at least on that fateful night before her exam. If Quinn had stayed in and studied, it would be because she was slightly different, at least that night. She would be such that she prioritized studying for her exam over a night out. But this applies to any decision we make in our lives. We decide to do what we do because of how we already are.

   But if what we do is because of the way we are, then in order to be responsible for our actions, we need to be the source of how we are, at least in the relevant mental respects (Strawson [1994] 2003, 219). There is the first premise. But here comes the rub: the way we are is a product of factors beyond our control such as the past and the laws of nature ([1994] 2003, 219; 222-223). The fact that Quinn is such that she prioritizes a night with friends over studying is due to her past and the relevant laws of nature. It is not up to her that she is the way she is. It is ultimately factors extending well beyond her, possibly all the way back to the initial conditions of the universe that account for why she is the way she is that night. And to the extent that this is compelling, the ultimate source of Quinn's decision to go out is not her. Rather, it is some condition of the universe external to her. And therefore, Quinn is not free.

   Once again, this is a difficult argument to respond to. You might note that "ultimate source" is ambiguous and needing further clarification. Some compatibilists have pointed this out and argued that once we start developing careful accounts of what it means to

be the source of our actions, we will see that the relevant notion of source-hood is compatible with determinism.

For example, while it may be true that no one is the ultimate cause of their actions in deterministic worlds precisely because the ultimate source of all actions will extend back to the initial conditions of the universe, we can still be a mediated source of our actions in the sense required for moral responsibility. Provided the actual source of our action involves a sophisticated enough set of capacities for it to make sense to view us as the source of our actions, we could still be the source of our actions, in the relevant sense (McKenna and Pereboom 2016, 154). After all, even if determinism is true, we still act for reasons. We still contemplate what to do and weigh reasons for and against various actions, and we still are concerned with whether or not the actions we are considering reflect our desires, our goals, our projects, and our plans. And you might think that if our actions stem from a history that includes us bringing all the features of our agency to bear upon the decision that is the proximal cause of our action, that this causal history is one in which we are the source of our actions in the way that is really relevant to identifying whether or not we are acting freely.

Others have noted that even if it is true that Quinn is not directly free in regard to the beliefs and desires that suggest she should go out with her friends rather than study (they are the product of factors beyond her control such as her upbringing, her environment, her genetics, or maybe even random luck), this need not imply that she lacks control as to whether or not she chooses to act upon them.[9] Perhaps it is the case that even though how we are may be due to factors beyond our control, nonetheless, we are still the source of what we do because it is still, even under determinism, up to us as to whether we choose to exercise control over our conduct.

Chapter Notes

9. For two attempts to respond to the ultimacy argument in this way, see: Mele, Alfred, 1995, *Autonomous Agents*, New York: Oxford University Press; and McKenna, Michael, 2008, "Ultimacy & Sweet Jane" in Nick Trakakis and Daniel Cohen, eds, *Essays on Free Will and Moral Responsibility*, Newcastle: Cambridge Scholars Publishing: 186-208.

Attribution

32. Reconciling Opposing Views

Determinism and Freedom

Determinism and free will are often thought to be in deep conflict. Whether or not this is true has a lot to do with what is meant by determinism and an account of what free will requires.

First of all, determinism is not the view that free actions are impossible. Rather, determinism is the view that at any one time, only one future is physically possible. To be a little more specific, determinism is the view that a complete description of the past along with a complete account of the relevant laws of nature logically entails all future events.[1]

Indeterminism is simply the denial of determinism. If determinism is incompatible with free will, it will be because free actions are only possible in worlds in which more than one future is physically possible at any one moment in time. While it might be true that free will requires indeterminism, it's not true merely by definition. A further argument is needed and this suggests that it is at least possible that people could sometimes exercise the control necessary for morally responsible action, even if we live in a deterministic world.

It is worth saying something about fatalism before we move on. It is really easy to mistake determinism for fatalism, and fatalism does seem to be in straightforward conflict with free will. Fatalism is the view that we are powerless to do anything other than what we actually do. If fatalism is true, then nothing that we try or think or intend or believe or decide has any causal effect or relevance as to what we actually end up doing.

But note that determinism need not entail fatalism. Determinism

is a claim about what is logically entailed by the rules/laws governing a world and the past of said world. It is not the claim that we lack the power to do other than what we actually were already going to do. Nor is it the view that we fail to be an important part of the causal story for why we do what we do. And this distinction may allow some room for freedom, even in deterministic worlds.

An example will be helpful here. We know that the boiling point for water is 100°C. Suppose we know in both a deterministic world and a fatalistic world that my pot of water will be boiling at 11:22am today. Determinism makes the claim that if I take a pot of water and I put it on my stove, and heat it to 100°C, it will boil. This is because the laws of nature (in this case, water that is heated to 100°C will boil) and the events of the past (I put a pot of water on a hot stove) bring about the boiling water. But fatalism makes a different claim. If my pot of water is fated to boil at 11:22am today, then no matter what I or anyone does, my pot of water will boil at exactly 11:22am today. I could try to empty the pot of water out at 11:21. I could try to take the pot as far away from a heating source as possible. Nonetheless, my pot of water will be boiling at 11:22 precisely because it was fated that this would happen. Under fatalism, the future is fixed or preordained, but this need not be the case in a deterministic world. Under determinism, the future is a certain way because of the past and the rules governing said world. If we know that a pot of water will boil at 11:22am in a deterministic world, it's because we know that the various causal conditions will hold in our world such that at 11:22 my pot of water will have been put on a heat source and brought to 100°C. Our deliberations, our choices, and our free actions may very well be part of the process that brings a pot of water to the boiling point in a deterministic world, whereas these are clearly irrelevant in fatalistic ones.

Three Views of Freedom

Most accounts of freedom fall into one of three camps. Some people take freedom to require merely the ability to "do what you want to do." For example, if you wanted to walk across the room, right now, and you also had the ability, right now, to walk across the room, you would be free as you could do exactly what you want to do. We will call this easy freedom.

Others view freedom on the infamous "Garden of Forking Paths" model. For these people, free action requires more than merely the ability to do what you want to do. It also requires that you have the ability to do otherwise than what you actually did. So, If Anya is free when she decides to take a sip from her coffee, on this view, it must be the case that Anya could have refrained from sipping her coffee. The key to freedom, then, is alternative possibilities and we will call this the alternative possibilities view of free action.

Finally, some people envision freedom as requiring, not alternative possibilities but the right kind of relationship between the antecedent sources of our actions and the actions that we actually perform. Sometimes this view is explained by saying that the free agent is the source, perhaps even the ultimate source of her action. We will call this kind of view a source view of freedom.

Now, the key question we want to focus on is whether or not any of these three models of freedom are compatible with determinism. It could turn out that all three kinds of freedom are ruled out by determinism, so that the only way freedom is possible is if determinism is false. If you believe that determinism rules out free action, you endorse a view called incompatibilism. But it could turn out that one or all three of these models of freedom are compatible with determinism. If you believe that free action is compatible with determinism, you are a compatibilist.

Let us consider compatibilist views of freedom and two of the most formidable challenges that compatibilists face: the consequence argument and the ultimacy argument.

Begin with easy freedom. Is easy freedom compatible with determinism? A group of philosophers called classic compatibilists certainly thought so.[2] They argued that free will requires merely the ability for an agent to act without external hindrance. Suppose, right now, you want to put your textbook down and grab a cup of coffee. Even if determinism is true, you probably, right now, can do exactly that. You can put your textbook down, walk to the nearest Starbucks, and buy an overpriced cup of coffee. Nothing is stopping you from doing what you want to do. Determinism does not seem to be posing any threat to your ability to do what you want to do right now. If you want to stop reading and grab a coffee, you can. But, by contrast, if someone had chained you to the chair you are sitting in, things would be a bit different. Even if you wanted to grab a cup of coffee, you would not be able to. You would lack the ability to do so. You would not be free to do what you want to do. This has nothing to do with determinism, of course. It is not the fact that you might be living in a deterministic world that is threatening your free will. It is that an external hindrance (the chains holding you to your chair) is stopping from you doing what you want to do. So, if what we mean by freedom is easy freedom, it looks like freedom really is compatible with determinism.

Easy freedom has run into some rather compelling opposition, and most philosophers today agree that a plausible account of easy freedom is not likely. But, by far, the most compelling challenge the view faces can be seen in the consequence argument.[3] The consequence argument is as follows:

1. If determinism is true, then all human actions are consequences of past events and the laws of nature.
2. No human can do other than they actually do

except by changing the laws of nature or changing the past.

3. No human can change the laws of nature or the past.

---

4. If determinism is true, no human has free will.

This is a powerful argument. It is very difficult to see where this argument goes wrong, if it goes wrong. The first premise is merely a restatement of determinism. The second premise ties the ability to do otherwise to the ability to change the past or the laws of nature, and the third premise points out the very reasonable assumption that humans are unable to modify the laws of nature or the past.

This argument effectively devastates easy freedom by proposing that we never act without external hindrances precisely because our actions are caused by past events and the laws of nature in such a way that we not able to contribute anything to the causal production of our actions. This argument also seems to pose a deeper problem for freedom in deterministic worlds. If this argument works, it establishes that, given determinism, we are powerless to do otherwise, and to the extent that freedom requires the ability to do otherwise, this argument seems to rule out free action. Note that if this argument works, it poses a challenge for both the easy and alternative possibilities view of free will.

How might someone respond to this argument? First, suppose you adopt an alternative possibilities view of freedom and believe that the ability to do otherwise is what is needed for genuine free will. What you would need to show is that alternative possibilities, properly understood, are not incompatible with determinism.

Perhaps you might argue that if we understand the ability to do otherwise properly we will see that we actually do have the ability to change the laws of nature or the past.

That might sound counterintuitive. How could it possibly be the case that a mere mortal could change the laws of nature or the past? Think back to Quinn's decision to spend the night before her exam out with friends instead of studying. When she shows up to her exam exhausted, and she starts blaming herself, she might say, "Why did I go out? That was dumb! I could have stayed home and studied." And she is sort of right that she could have stayed home. She had the general ability to stay home and study. It is just that if she had stayed home and studied the past would be slightly different or the laws of nature would be slightly different. What this points to is that there might be a way of cashing out the ability to do otherwise that is compatible with determinism and does allow for an agent to kind of change the past or even the laws of nature.[4]

But suppose we grant that the consequence argument demonstrates that determinism really does rule out alternative possibilities. Does that mean we must abandon the alternative possibilities view of freedom? Well, not necessarily. You could instead argue that free will is possible, provided determinism is false.[5] That is a big if, of course, but maybe determinism will turn out to be false.

What if determinism turns out to be true? Should we give up, then, and concede that there is no free will? Well, that might be too quick. A second response to the consequence argument is available. All you need to do is deny that freedom requires the ability to do otherwise.

In 1969, Harry Frankfurt proposed an influential thought experiment that demonstrated that free will might not require alternative possibilities at all (Frankfurt [1969] 1988). If he's right about this, then the consequence argument, while compelling, does not demonstrate that no one lacks free will in deterministic worlds, because free will does not require the ability to do otherwise. It merely requires that agents be the source of their actions in the

right kind of way. But we're getting ahead of ourselves. Here is a simplified paraphrase of Frankfurt's case:

> Black wants Jones to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid unnecessary work. So he waits until Jones is about to make up his mind what to do, and he does nothing unless it is clear to him (Black is an excellent judge of such things) that Jones is going to decide not to do what Black wants him to do. If it does become clear that Jones is going to decide to do something other than what Black wanted him to do, Black will intervene, and ensure that Jones decides to do, and does do, exactly what Black wanted him to do. Whatever Jones' initial preferences and inclinations, then, Black will have his way. As it turns out, Jones decides, on his own, to do the action that Black wanted him to perform. So, even though Black was entirely prepared to intervene, and could have intervened, to guarantee that Jones would perform the action, Black never actually has to intervene because Jones decided, for reasons of his own, to perform the exact action that Black wanted him to perform. (Frankfurt [1969] 1988, 6-7)

Now, what is going on here? Jones is overdetermined to perform a specific act. No matter what happens, no matter what Jones initially decides or wants to do, he is going to perform the action Black wants him to perform. He absolutely cannot do otherwise. But note that there seems to be a crucial difference between the case in which Jones decides on his own and for his own reasons to perform the action Black wanted him to perform and the case in which Jones would have refrained from performing the action were it not for Black intervening to force him to perform the action. In the first case, Jones is the source of his action. It the thing he decided to do and he does it for his own reasons. But in the second case, Jones is not the source of his actions. Black is. This distinction, thought Frankfurt, should be at the heart of discussions of free will and

moral responsibility. The control required for moral responsibility is not the ability to do otherwise (Frankfurt [1969] 1988, 9-10).

If alternative possibilities are not what free will requires, what kind of control is needed for free action? Here we have the third view of freedom we started with: free will as the ability to be the source of your actions in the right kind of way. Source compatibilists argue that this ability is not threatened by determinism, and building off of Frankfurt's insight, have gone on to develop nuanced, often radically divergent source accounts of freedom.[6] Should we conclude, then, that provided freedom does not require alternative possibilities that it is compatible with determinism?[7] Again, that would be too quick. Source compatibilists have reason to be particularly worried about an argument developed by Galen Strawson called the ultimacy argument (Strawson [1994] 2003, 212-228).

Rather than trying to establish that determinism rules out alternative possibilities, Strawson tried to show that determinism rules out the possibility of being the ultimate source of your actions. While this is a problem for anyone who tries to establish that free will is compatible with determinism, it is particularly worrying for source compatibilists as they've tied freedom to an agent's ability to be source of its actions. Here is the argument:

1. A person acts of her own free will only if she is the act's ultimate source.
2. If determinism is true, no one is the ultimate source of her actions.

---

3. Therefore, if determinism is true, no one acts of

her own free will. (McKenna and Pereboom 2016, 148)[8]

This argument requires some unpacking. First of all, Strawson argues that for any given situation, we do what we do because of the way we are ([1994] 2003, 219). When Quinn decides to go out with her friends rather than study, she does so because of the way she is. She prioritizes a night with her friends over studying, at least on that fateful night before her exam. If Quinn had stayed in and studied, it would be because she was slightly different, at least that night. She would be such that she prioritized studying for her exam over a night out. But this applies to any decision we make in our lives. We decide to do what we do because of how we already are.

But if what we do is because of the way we are, then in order to be responsible for our actions, we need to be the source of how we are, at least in the relevant mental respects (Strawson [1994] 2003, 219). There is the first premise. But here comes the rub: the way we are is a product of factors beyond our control such as the past and the laws of nature ([1994] 2003, 219; 222-223). The fact that Quinn is such that she prioritizes a night with friends over studying is due to her past and the relevant laws of nature. It is not up to her that she is the way she is. It is ultimately factors extending well beyond her, possibly all the way back to the initial conditions of the universe that account for why she is the way she is that night. And to the extent that this is compelling, the ultimate source of Quinn's decision to go out is not her. Rather, it is some condition of the universe external to her. And therefore, Quinn is not free.

Once again, this is a difficult argument to respond to. You might note that "ultimate source" is ambiguous and needing further clarification. Some compatibilists have pointed this out and argued that once we start developing careful accounts of what it means to

be the source of our actions, we will see that the relevant notion of source-hood is compatible with determinism.

For example, while it may be true that no one is the ultimate cause of their actions in deterministic worlds precisely because the ultimate source of all actions will extend back to the initial conditions of the universe, we can still be a mediated source of our actions in the sense required for moral responsibility. Provided the actual source of our action involves a sophisticated enough set of capacities for it to make sense to view us as the source of our actions, we could still be the source of our actions, in the relevant sense (McKenna and Pereboom 2016, 154). After all, even if determinism is true, we still act for reasons. We still contemplate what to do and weigh reasons for and against various actions, and we still are concerned with whether or not the actions we are considering reflect our desires, our goals, our projects, and our plans. And you might think that if our actions stem from a history that includes us bringing all the features of our agency to bear upon the decision that is the proximal cause of our action, that this causal history is one in which we are the source of our actions in the way that is really relevant to identifying whether or not we are acting freely.

Others have noted that even if it is true that Quinn is not directly free in regard to the beliefs and desires that suggest she should go out with her friends rather than study (they are the product of factors beyond her control such as her upbringing, her environment, her genetics, or maybe even random luck), this need not imply that she lacks control as to whether or not she chooses to act upon them.[9] Perhaps it is the case that even though how we are may be due to factors beyond our control, nonetheless, we are still the source of what we do because it is still, even under determinism, up to us as to whether we choose to exercise control over our conduct.

Free Will and the Sciences

Many challenges to free will come, not from philosophy, but from the sciences. There are two main scientific arguments against free will, one coming from neuroscience and one coming from the social sciences. The concern coming from research in the neurosciences is that some empirical results suggest that all our choices are the result of unconscious brain processes, and to the extent choices must be consciously made to be free choices, it seems that we never make a conscious free choice.

The classic studies motivating a picture of human action in which unconscious brain processes are doing the bulk of the causal work for action were conducted by Benjamin Libet. Libet's experiments involved subjects being asked to flex their wrists whenever they felt the urge to do so. Subjects were asked to note the location of a clock hand on a modified clock when they became aware of the urge to act. While doing this their brain activity was being scanned using EEG technology. What Libet noted is that around 550 milliseconds before a subject acted, a readiness potential (increased brain activity) would be measured by the EEG technology. But subjects were reporting awareness of an urge to flex their wrist around 200 milliseconds before they acted (Libet 1985).

This painted a strange picture of human action. If conscious intentions were the cause of our actions, you may expect to see a causal story in which the conscious awareness of an urge to flex your wrist shows up first, then a ramping up of brain activity, and finally an action. But Libet's studies showed a causal story in which an action starts with unconscious brain activity, the subject later becomes consciously aware that they are about to act, and then the action happens. The conscious awareness of action seemed to be a byproduct of the actual unconscious process that was causing the action. It was not the cause of the action itself. And this result suggests that unconscious brain processes, not conscious ones, are the real causes of our actions. To the extent that free action requires

our conscious decisions to be the initiating causes of our actions, it looks like we may never act freely.

While this research is intriguing, it probably does not establish that we are not free. Alfred Mele is a philosopher who has been heavily critical of these studies. He raises three main objections to the conclusions drawn from these arguments.

First, Mele points out that self-reports are notoriously unreliable (2009, 60-64). Conscious perception takes time, and we are talking about milliseconds. The actual location of the clock hand is probably much closer to 550 milliseconds when the agent "intends" or has the "urge" to act than it is to 200 milliseconds. So, there's some concerns about experimental design here.

Second, an assumption behind these experiments is that what is going on at 550 milliseconds is that a decision is being made to flex the wrist (Mele 2014, 11). We might challenge this assumption. Libet ran some variants of his experiment in which he asked subjects to prepare to flex their wrist but to stop themselves from doing so. So, basically, subjects simply sat there in the chair and did nothing. Libet interpreted the results of these experiments as showing that we might not have a free will, but we certainly have a "free won't" because we seem capable of consciously vetoing or stopping an action, even if that action might be initiated by unconscious processes (2014, 12-13). Mele points out that what might be going on in these scenarios is that the real intention to act or not act is what happens consciously at 200 milliseconds, and if so, there is little reason to think these experiments are demonstrating that we lack free will (2014, 13).

Finally, Mele notes that while it may be the case that some of our decisions and actions look like the wrist-flicking actions Libet was studying, it is doubtful that all or even most of our decisions are like this (2014, 15). When we think about free will, we rarely think of actions like wrist-flicking. Free actions are typically much more complex and they are often the kind of thing where the decision to do something extends across time. For example, your decision about what to major in at college or even where to study was

probably made over a period of months, even years. And that decision probably involved periods of both conscious and unconscious cognition. Why think that a free choice cannot involve some components that are unconscious?

A separate line of attack on free will comes from the situationist literature in the social sciences (particularly social psychology). There is a growing body of research suggesting that situational and environmental factors profoundly influence human behavior, perhaps in ways that undermine free will (Mele 2014, 72).

Many of the experiments in the situationist literature are among the most vivid and disturbing in all of social psychology. Stanley Milgram, for example, conducted a series of experiments on obedience in which ordinary people were asked to administer potentially lethal voltages of electricity to an innocent subject in order to advance scientific research, and the vast majority of people did so![10] And in Milgram's experiments, what affected whether or not subjects were willing to administer the shocks were minor, seemingly insignificant environmental factors such as whether the person running the experiment looked professional or not (Milgram 1963).

What experiments like Milgram's obedience experiments might show is that it is our situations, our environments that are the real causes of our actions, not our conscious, reflective choices. And this may pose a threat to free will. Should we take this kind of research as threatening freedom?

Many philosophers would resist concluding that free will does not exist on the basis of these kinds of experiments. Typically, not everyone who takes part in situationist studies is unable to resist the situational influences they are subject to. And it appears to be the case that when we are aware of situational influences, we are more likely to resist them. Perhaps the right way to think about this research is that there all sorts of situations that can influence us in ways that we may not consciously endorse, but that nonetheless, we are still capable of avoiding these effects when we are actively trying to do so. For example, the brain sciences have made many of

us vividly aware of a whole host of cognitive biases and situational influences that humans are typically subject to and yet, when we are aware of these influences, we are less susceptible to them. The more modest conclusion to draw here is not that we lack free will, but that exercising control over our actions is much more difficult than many of us believe it to be. We are certainly influenced by the world we are a part of, but to be influenced by the world is different from being determined by it, and this may allow us to, at least sometimes, exercise some control over the actions we perform.

No one knows yet whether or not humans sometimes exercise the control over their actions required for moral responsibility. And so I leave it to you, dear reader: Are you free?

Chapter Notes

1. I have hidden some complexity here. I have defined determinism in terms of logical entailment. Sometimes people talk about determinism as a causal relationship. For our purposes, this distinction is not relevant, and if it is easier for you to make sense of determinism by thinking of the past and the laws of nature causing all future events, that is perfectly acceptable to do.
2. Two of the more well-known classic compatibilists include Thomas Hobbes and David Hume. See: Hobbes, Thomas, (1651) 1994, *Leviathan*, ed. Edwin Curley, Canada: Hackett Publishing Company; and Hume, David, (1739) 1978, A *Treatise of Human Nature*, Oxford: Oxford University Press.
3. For an earlier version of this argument see: Ginet, Carl, 1966, "Might We Have No Choice?" in *Freedom and Determinism*, ed. Keith Lehrer, 87-104, Random House.
4. For two notable attempts to respond to the consequence argument by claiming that humans can change the past or the laws of nature see: Fischer, John Martin, 1994, *The Metaphysics*

of Free Will, Oxford: Blackwell Publishers; and Lewis, David, 1981, "Are We Free to Break the Laws?" *Theoria* 47: 113-21.

5. Many philosophers try to develop views of freedom on the assumption that determinism is incompatible with free action. The view that freedom is possible, provided determinism is false is called Libertarianism. For more on Libertarian views of freedom, see: Clarke, Randolph and Justin Capes, 2017, "Incompatibilist (Nondeterministic) Theories of Free Will," *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/incompatibilism-theories/.

6. For elaboration on recent compatibilist views of freedom, see McKenna, Michael and D. Justin Coates, 2015, "Compatibilism," *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/compatibilism/.

7. You might be unimpressed by the way source compatibilists understand the ability to be the source of your actions. For example, you might that what it means to be the source of your actions is to be the ultimate cause of your actions. Or maybe you think that to genuinely be the source of your actions you need to be the agent-cause of your actions. Those are both reasonable positions to adopt. Typically, people who understand free will as requiring either of these abilities believe that free will is incompatible with determinism. That said, there are many Libertarian views of free will that try to develop a plausible account of agent causation. These views are called Agent-Causal Libertarianism. See: Clarke, Randolph and Justin Capes, 2017, "Incompatibilist (Nondeterministic) Theories of Free Will," *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/incompatibilism-theories/.

8. As with most philosophical arguments, the ultimacy argument has been formulated in a number of different ways. In Galen Strawson's original paper he gives three different versions of the argument, one of which has eight premises and one that

has ten premises. A full treatment of either of those versions of this argument would require more time and space than we have available here. I have chosen to use the McKenna/Pereboom formulation of the argument due its simplicity and their clear presentation of the central issues raised by the argument.

9. For two attempts to respond to the ultimacy argument in this way, see: Mele, Alfred, 1995, *Autonomous Agents*, New York: Oxford University Press; and McKenna, Michael, 2008, "Ultimacy & Sweet Jane" in Nick Trakakis and Daniel Cohen, eds, *Essays on Free Will and Moral Responsibility*, Newcastle: Cambridge Scholars Publishing: 186-208.

10. Fortunately, no real shocks were administered. The subjects merely believed they were doing so.

References

Frankfurt, Harry. (1969) 1988. "Alternative Possibilities and moral responsibility." In *The Importance of What We Care About: Philosophical Essays*, 10th ed. New York: Cambridge University Press.

Libet, Benjamin. 1985. "Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action." *Behavioral and Brain Sciences* 8: 529-566.

McKenna, Michael and Derk Pereboom. 2016. *Free Will: A Contemporary Introduction.* New York: Routledge.

Mele, Alfred. 2014. *Free: Why Science Hasn't Disproved Free Will.* Oxford: Oxford University Press.

Mele, Alfred. 2009. *Effective Intentions: The Power of Conscious Will.* Oxford: Oxford University Press.

Milgram, Stanley. 1963. "Behavioral Study of Obedience." *The Journal of Abnormal and Social Psychology* 67: 371-378.

Strawson, Galen. (1994) 2003. "The Impossibility of Moral

Responsibility." In *Free Will*, 2nd ed. Edited by Gary Watson, 212-228. Oxford: Oxford University Press.

van Inwagen, Peter. 1983. *An Essay on Free Will.* Oxford: Clarendon Press.

Further Reading

Deery, Oisin and Paul Russell, eds. 2013. *The Philosophy of Free Will: Essential Readings from the Contemporary Debates*. New York: Oxford University Press.

Mele, Alfred. 2006. *Free Will and Luck.* New York: Oxford University Press.

Attribution

This section is composed of text taken from *Chapter 8 Freedom of the Will* created by Daniel Haas in *Introduction to Philosophy: Philosophy of Mind*, edited by Heather Salazar and Christina Hendricks, and produced with support from the Rebus Community. The original is freely available under the terms of the CC BY 4.0 license at https://press.rebus.community/intro-to-phil-of-mind/. The material is presented in its original form, with the exception of the removal of introductory material *Introduction: Are We Free?*

33. Determinism

Determinism and Freedom

Determinism and free will are often thought to be in deep conflict. Whether or not this is true has a lot to do with what is meant by determinism and an account of what free will requires.

First of all, determinism is not the view that free actions are impossible. Rather, determinism is the view that at any one time, only one future is physically possible. To be a little more specific, determinism is the view that a complete description of the past along with a complete account of the relevant laws of nature logically entails all future events.[1]

Indeterminism is simply the denial of determinism. If determinism is incompatible with free will, it will be because free actions are only possible in worlds in which more than one future is physically possible at any one moment in time. While it might be true that free will requires indeterminism, it's not true merely by definition. A further argument is needed and this suggests that it is at least possible that people could sometimes exercise the control necessary for morally responsible action, even if we live in a deterministic world.

It is worth saying something about fatalism before we move on. It is really easy to mistake determinism for fatalism, and fatalism does seem to be in straightforward conflict with free will. Fatalism is the view that we are powerless to do anything other than what we actually do. If fatalism is true, then nothing that we try or think or intend or believe or decide has any causal effect or relevance as to what we actually end up doing.

But note that determinism need not entail fatalism. Determinism is a claim about what is logically entailed by the rules/laws governing a world and the past of said world. It is not the claim that

we lack the power to do other than what we actually were already going to do. Nor is it the view that we fail to be an important part of the causal story for why we do what we do. And this distinction may allow some room for freedom, even in deterministic worlds.

An example will be helpful here. We know that the boiling point for water is 100°C. Suppose we know in both a deterministic world and a fatalistic world that my pot of water will be boiling at 11:22am today. Determinism makes the claim that if I take a pot of water and I put it on my stove, and heat it to 100°C, it will boil. This is because the laws of nature (in this case, water that is heated to 100°C will boil) and the events of the past (I put a pot of water on a hot stove) bring about the boiling water. But fatalism makes a different claim. If my pot of water is fated to boil at 11:22am today, then no matter what I or anyone does, my pot of water will boil at exactly 11:22am today. I could try to empty the pot of water out at 11:21. I could try to take the pot as far away from a heating source as possible. Nonetheless, my pot of water will be boiling at 11:22 precisely because it was fated that this would happen. Under fatalism, the future is fixed or preordained, but this need not be the case in a deterministic world. Under determinism, the future is a certain way because of the past and the rules governing said world. If we know that a pot of water will boil at 11:22am in a deterministic world, it's because we know that the various causal conditions will hold in our world such that at 11:22 my pot of water will have been put on a heat source and brought to 100°C. Our deliberations, our choices, and our free actions may very well be part of the process that brings a pot of water to the boiling point in a deterministic world, whereas these are clearly irrelevant in fatalistic ones.

Chapter Notes

1. I have hidden some complexity here. I have defined determinism in terms of logical entailment. Sometimes people

talk about determinism as a causal relationship. For our purposes, this distinction is not relevant, and if it is easier for you to make sense of determinism by thinking of the past and the laws of nature causing all future events, that is perfectly acceptable to do.

Attribution

This section is composed of text taken from *Chapter 8 Freedom of the Will* created by Daniel Haas in *Introduction to Philosophy: Philosophy of Mind*, edited by Heather Salazar and Christina Hendricks, and produced with support from the Rebus Community. The original is freely available under the terms of the CC BY 4.0 license at https://press.rebus.community/intro-to-phil-of-mind/. The material is presented in its original form, with the exception of the removal of introductory material *Introduction: Are We Free?*

34. Attribution

This chapter is composed of text taken from *Chapter 8 Freedom of the Will* created by Daniel Haas in *Introduction to Philosophy: Philosophy of Mind*, edited by Heather Salazar and Christina Hendricks, and produced with support from the Rebus Community. The original is freely available under the terms of the CC BY 4.0 license at https://press.rebus.community/intro-to-phil-of-mind/. The material is presented in its original form, with the exception of the removal of introductory material *Introduction: Are We Free?*

INTRODUCING MORAL THEORIES

35. Right Action

Right Action

Our focus in this chapter will be normative ethics. Normative ethical principles aren't intended to describe how things are, how people think or how they behave. Normative ethics is concerned how we should be motivated and how we should act. Our project here is to think critically about which normative ethical principles do the best job of explaining our assorted moral intuitions about the broadest range of possible cases. We will start with Utilitarianism, a view of right action based on the idea that happiness has fundamental value. We'll then examine Kant's ethics of respect for persons. On this view persons have intrinsic moral worth, and ethics is concerned with what respecting the value of persons requires of us.

Both Utilitarianism and Kant's ethics of respect for persons can be understood as aiming to formulate action-guiding normative ethical principles. Later in the chapter we will consider approaches to normative ethics that are not so concerned with identifying exceptionless "laws" of right action. Our understanding of right action doesn't have to be expressible in terms of strict rules. Feminist ethics finds value in caring relationships. But taking relationships to be good doesn't directly lead to specific rules for action as Utilitarianism might. Environmental ethicists have advanced various proposals for expanding the realm of moral relevance to include other species or systems of life as a whole. This is not to deny that people matter morally, but many environmental ethicists deny that people are all that matter. Accounting for the value of non-persons in addition to persons is likely to frustrate attempts to

characterize right action in terms of simple formulas or "moral laws."

At the end of this chapter we will consider a pluralistic approach to understanding ethical motivation and action. The suggestion here will be that a substantive realist approach to normative ethics doesn't require reducing all ethical value to one fundamental kind. Such a pluralistic account of ethical value undermines the quest for simple exceptionless or absolute moral principles. But it also suggests that substantive realist normative ethics doesn't require these either.

Utilitarianism

Utilitarianism is based on the idea that happiness is good. Utilitarian thinkers have traditionally understood happiness in terms of pleasure and the absence of pain. Utilitarianism's best known advocate, John Stuart Mill, characterizes Utilitarianism as the view that "an action is right insofar as it tends to produce pleasure and the absence of pain." If happiness, conceived of as pleasure and the absence of pain, is the one thing that has value, then this criterion of right action should seem to follow straightforwardly.

In any given scenario, every possible course of action will have a utility. The utility of an action is the net total of pleasure caused by the action minus any pain caused by that action. In calculating the utility of an action we are to consider all of the effects of the action, both long run and short run. Given the utilities of all available courses of action, Utilitarianism says that the correct course of action is the one that has the greatest utility. So an action is right if it produces the greatest net total of pleasure over pain of any available alternative action. Note that sometimes no possible course of action will produce more pleasure than

pain. This is not a problem for Utilitarianism as we've formulated it. Utilitarianism will simply require us to pursue the lesser evil. The action with the highest utility can still have negative utility.

Utilitarianism places no privileged status on the happiness of the actor. It's happiness that matters, not just your happiness. So Utilitarianism can call for great personal sacrifice. The happiness of my child over the course of his lifetime might require great personal sacrifice on my part over the course of his first few decades. Utilitarianism says the sacrifice should be made given that the utility at stake for my child is greater than the utility at stake in my child-rearing sacrifices.

Likewise, Utilitarianism places no privileged status on the immediate, as opposed to the long term, effects of the action. An action's utility is the net amount of pleasure or pain that is experienced as a result of the action over the long run. So, while it might maximize a small child's pleasure in the short run to be given ice cream whenever he wants it, the long run utility of this might not be so good given the habits formed and the health consequences of an over- indulged sweet tooth.

There is an obvious concern to address at this point. We often don't know what the long-run consequences of our actions will be, and even in the short run we are often uncertain about just how much pleasure and pain will be caused for the various parties affected. So we might not be able to calculate the utilities of alternative actions to figure out which action will have the highest utility. These are practical problems for applying utilitarian theory. But while it might be difficult to tell on a case by case basis just which course of action will maximize utility, this is not a problem for Utilitarianism as a normative ethical theory. As a normative ethical theory, Utilitarianism is aimed at identifying the standard for right action, not telling when a particular action meets that standard. Setting the standard for right action and figuring out how to meet that standard are two different projects.

When we speak of utility as pleasure and the absence of pain,

we need to take "pleasure" and "pain" in the broadest sense possible. There are social, intellectual, and aesthetic pleasures to consider, as well as sensual pleasures. Recognizing this is important to answering what Mill calls the "doctrine of swine" objection to Utilitarianism. This objection takes Utilitarianism to be unfit for humans because it recognizes no higher purpose to life than the mere pursuit of pleasure. The objector takes people to have more noble ends to pursue than mere pleasure. According to this objection, Utilitarianism is a view of the good that is fit only for swine. Mill responds that it is the person who raises this objection who portrays human nature in a degrading light, not the utilitarian theory of right action. People are capable of pleasures beyond mere sensual indulgences and the utilitarian theory concerns these as well. Mill then argues that social and intellectual pleasures are of an intrinsically higher quality than sensual pleasure.

We find a more significant objection to Utilitarian moral theory in the following sort of case: Consider Bob, who goes to the doctor for a checkup. His doctor finds that Bob is in perfect health. And his doctor also finds that Bob is biologically compatible with six other patients she has who are all dying of various sorts of organ failure. Let's assume that if Bob lives out his days he will live a typically good life, one that is pleasant to Bob and also brings happiness to his friends and family. But we will assume that Bob will not discover a cure for AIDS or bring about world peace. And let us make similar assumptions about the six people suffering from organ failure. According to simple Act Utilitarianism, it looks like the right thing for Bob's doctor to do is to kill Bob and harvest his organs for the benefit of the six patients who will otherwise die. But intuitively, this would be quite wrong. Act Utilitarianism gets the wrong result in this sort of case. This case seems to provide a clear counterexample to simple Act Utilitarianism. This looks like a bit of evidence that calls for a change in theory. But perhaps that change can

be a modification of utilitarian thinking rather than a complete rejection of it.

One move open to the utilitarian is to evaluate rules for acting rather than individual actions. A version of Rule Utilitarianism might say that the right action is the action that follows the rule which, in general, will produce the highest utility. A rule that tells doctors to kill their patients when others require their organs would not have very high utility in general. People would avoid their doctors and illness would go untreated were such a rule in effect. Rather, the rule that doctors should do no harm to their patients would have much higher utility in general. So the move to Rule Utilitarianism seems to avoid the difficulty we found with Act Utilitarianism. Or at least it seems to when we consider just these two rules.

But here is a rule that would have even higher utility than the rule that doctors should never harm their patients: doctors should never harm their patients except when doing so would maximize utility. Now suppose that doctors ordinarily refrain from harming their patients and as a result people trust their doctors. But in Bob's case, his doctor realizes that she can maximize utility by killing Bob and distributing his organs. She can do this in a way that no one will ever discover, so her harming Bob in this special case will not undermine people's faith in the medical system. The possibility of rules with "except when utility is maximized" clauses renders Rule Utilitarianism vulnerable to the same kinds of counterexamples we found for Act Utilitarianism. In effect, Rule Utilitarianism collapses back into Act Utilitarianism.

In order to deal with the original problem of Bob and his vital organs, the advocate of Rule Utilitarianism must find a principled way to exclude certain sorts of utility maximizing rules. I won't pursue this matter on behalf of the utilitarian. Rather, I want to consider further just how simple Act Utilitarianism goes wrong in Bob's case. Utilitarianism evaluates the goodness of actions in terms of their

consequences. For this reason, Utilitarianism is often referred to as a *consequentialist* theory. Utilitarian considerations of good consequences seem to leave out something that is ethically important. Specifically, in this case, it leaves out a proper regard for Bob as person with a will of his own. What makes Bob's case a problem case is something other than consequences, namely, his status as a person and the sort of regard this merits. This problem case for utilitarian moral theory seems to point towards the need for a theory based on the value of things other than an action's consequences. Such non-consequentialist ethical theory is called *deontological* ethical theory. The best known deontological theory is the ethics of respect for persons. And this will be our next topic.

Here is a link to John Stuart Mill's essay *Utilitarianism: http://www.gutenberg.org/files/11224/11224-h/11224-h.htm*

36. Kantian Ethics

Respect for Persons: Kant's Moral Theory

Like Utilitarianism, Imannual Kant's moral theory is grounded in a theory of intrinsic value. But where the utilitarian takes happiness, conceived of as pleasure and the absence of pain to be what has intrinsic value, Kant takes the only thing to have moral worth for its own sake to be the capacity for good will we find in persons. Persons, conceived of as autonomous rational moral agents, are beings that have intrinsic moral worth and hence beings that deserve moral respect.

The opening passage of Immanuel Kant's *Groundwork for a Metaphysic of Morals* proclaims that "it is impossible to conceive of anything in the world, or indeed beyond it, that can be understood as good without qualification except for a good will." This is a clear and elegant statement of the theory of value that serves as the basis for Kant's ethical theory of respect for persons. The one thing that has intrinsic value, for Kant, is the autonomous good will of a

person. That said, Kant does not understand the expression "good will" in the everyday sense. In everyday discourse we might speak of someone being a person of good will if they want to do good things. We take the philanthropist's desire to give to the less fortunate to be an example of good will in this everyday sense. On Kant's view, the person of good will wills good things, but out of a sense of moral duty, not just inclination. Naturally generous philanthropists do not demonstrate their good will through their giving according to Kant, but selfish greedy persons do show their good will when they give to the poor out of a recognition of their moral duty to do so even though they'd really rather not. So it is our ability to recognize a moral duty

and will to act in accordance with it that makes persons beings that have dignity and are therefore worthy of moral regard. On Kant's view, our free will, our moral autonomy, is our capacity to act according to duty as opposed to being a slave to our desires or inclinations. So free will, in the sense that is associated with moral responsibility, doesn't mean being free to do as you please without consequence. Rather, freedom comes with moral responsibility for the intentions we act on.

So, understanding the good will as the capacity to will and act out of duty or respect for moral law, we can see having this capacity as part of having a rational, autonomous will. As persons, we have a free or autonomous will in our capacity to weigh our desires against each other and against the rational constraints of morality and reach our own determination of the will. We are

the originators and authors of the principles we act on. On Kant's view, our free will, our moral autonomy, is our capacity to act according to duty as opposed to being a slave to our desires or inclinations. So free will, in the sense that is associated with moral responsibility, doesn't mean being free to do as you please without consequence. Rather, freedom comes with moral responsibility for the intentions we act on. Having an autonomous good will with the capacity to act from moral duty is central to being a person in the moral sense and it is the basis, the metaphysical grounding, for an ethics of respect for persons. Now what it is to respect a person merits some further analysis.

Kant calls his fundamental moral principle the Categorical Imperative. An imperative is a command. The notion of a Categorical Imperative can be understood in contrast to that of a hypothetical imperative. A hypothetical imperative tells you what to do in order to achieve some goal. For instance, "if you want to get a good grade in calculus, work the assignments regularly." This claim tells you what to do in order to get a good grade in calculus. But it doesn't tell you what to do if you don't care about getting a good grade. What is distinctive

about a Categorical Imperative is that it tells you how to act regardless of what end or goal you might desire. Kant holds that if there is a fundamental law of morality, it is a Categorical Imperative. Taking the fundamental principle of morality to be a Categorical Imperative implies that moral reasons override other sorts of reasons. You might, for instance, think you have a self-interested reason to cheat on exam. But if morality is grounded in a Categorical Imperative, then your moral reason against cheating overrides your self-interested reason for cheating. If we think considerations of moral obligation trump self-interested considerations, Kant's idea that the fundamental law of morality is a Categorical Imperative accounts for this nicely.

Here are two formulations of Kant's Categorical Imperative:

CIa: Always treat persons (including yourself) as ends in themselves, never merely as a means to an end.

CIb: Act only on that maxim that you can consistently will to be a universal law.

Kant takes these formulations to be different ways of expressing the same underlying principle of respect for persons. They certainly don't appear to be synonymous. But we might take them to express the same thing in that each formulation would guide one to act in the same way.

The formulation (CIa), tells us to treat individuals as ends in themselves. That is just to say that persons should be treated as beings that have intrinsic value. To say that persons have intrinsic value is to say that they have value independent of their usefulness for this or that purpose. (CIa) does not say that you can never use a person for your own purposes. But it tells us we should never use a person merely as a means to your own ends. What is the difference? We treat people as a means to our own ends in ways that are not morally problematic quite often. When I go to the post office, I treat the clerk as a means to my end of sending a letter. But I do not treat that person merely as

a means to an end. I pursue my end of sending a letter through my interaction

with the clerk only with the understanding that the clerk is acting autonomously in serving me. My interaction with the clerk is morally acceptable so long as the clerk is serving me voluntarily, or acting autonomously for his own reasons. By contrast, we use people merely as a means to an end if we force them to do our will, or if we deceive them into doing our will. Coercion and deception are paradigm violations of the Categorical Imperative. In coercing or deceiving another person, we disrupt his or her autonomy and his or her will. This is what the Categorical Imperative forbids. Respecting persons requires refraining from violating their autonomy.

Now let's consider the second formulation CIb. This version, known as the formula of the universal law, tells us to "act only on that maxim that you could consistently will to be a universal law." The maxim of our action is the subjective principle that determines our will. We act for our own reasons. Different intentions might lead to similar actions. When I want to make myself a bit more presentable, I shave and shower. My son might perform the same action for a different reason (to get his mom off his back, for instance). We can identify different maxims in terms of these different reasons or intentions. For Kant, intentions matter. He evaluates the moral status of actions not according to the action itself or according to its consequences, but according to the maxim of the action. The moral status of an action is determined by the actor's intentions or reasons for acting.

According to the formula of the universal law, what makes an action morally acceptable is that its maxim is universalizable. That is, morally permissible action is action that is motivated by an intention that we can rationally will that others act on similarly. A morally prohibited action is just one where we can't rationally will that our maxim is universally followed. Deception and coercion are both paradigm cases of acting wrongly

according to Kant. In both cases, our maxim involves violating the autonomy of another rational being and this is something that we, as rationally autonomous beings ourselves, could not consistently will to be a universal law.

According to Kant, there is a contradiction involved in a rational autonomous being willing that autonomy be universally coercively or deceptively violated. This would involve a rational autonomous being willing the violation of its own rational autonomy. Acting out of moral duty is a matter of acting only on maxims that we can rationally will others act on as well. The person of good will recognizes the humanity of others by not making any special exception for herself even when her interests or inclination would be served by doing so.

There is no higher moral authority than the rational autonomous person, according to Kant. Morality is not a matter of following rules laid down by some higher authority. It is rather a matter of writing rules for ourselves that are compatible with the rational autonomous nature we share with other persons. We show respect for others through restraining our own will in ways that demonstrate our recognition of them as moral equals.

Primary Source Reading:

Kant's *Groundwork for a Metaphysic of Morals* can be found here: http://www.earlymoderntext s.com/assets/pdfs/ kant1785.pdf

Ethical Pluralism

In ethical theory, we can understand pluralism as the view that there is a plurality of fundamentally good things. Traditionally, ethicists have tried to analyze right and wrong action in terms

of a single fundamental underlying kind of value. We can call this kind of approach *ethical monism.* For utilitarians that single value is happiness, for Kantian respect for persons theorists, it is the value of the person. Ethical Pluralism allows that there may be multiple kinds of fundamental and irreducible value in the world. Happiness and respect for persons might be among these, but there may be others yet. Here I'll explain how pluralism so understood differs from Moral Relativism and how it is better suited than relativism and monist ethical theories to the goals of social justice sought by pluralism in a broader sense of valuing diversity.

Recall that according to Moral Relativism, what makes something right relative to a group is just that it is deemed to be right by that group. This is a pretty loose characterization of the view. We could get a bit more specific by asking just what the relevant groups are. We would also want to ask who gets to decide for that group, because according to Moral Relativism and other conventionalist views of morality (like Divine Command Theory) right and wrong, good and bad, are ultimately questions of authority.

Views that take morality to be matter of authority, whether it's God's, the culture's collectively, the king's or the chess club's authority, all suffer the same basic defect. They render right and wrong entirely arbitrary. If someone or some group gets to decide what's right and wrong, then anything can be right or wrong. According to Cultural Moral Relativism, whatever a culture deems to be morally right is right relative to it. So, if our culture says that homophobia, sexism, and racism are fine, then they are what is right relative to our culture and that's the end of it. If some people don't like it, that's just too bad. Moral Relativism denies them any objective standpoint from which to complain or any possibility of providing reasons for changing things. Complaints about the oppressiveness of the dominant group amount to nothing more than the whininess of losers. The group that dominates is perfectly well within

its rights to do so. This hardly sounds like a plausible account of social justice. But it is straightforwardly entailed by Moral Relativism and that's exactly why Moral Relativism is an awful ethical theory. This much is just a bit of review from the last chapter. But bear this in mind for the purpose of recognizing how Ethical Pluralism avoids this defect. For according to Ethical Pluralism the fundamental ethical values are real. The importance of happiness comes with the existence of pleasure. The value of respect for persons comes with the existence of persons. This doesn't depend on the whim or say so of any authority.

Suppose morality doesn't depend on the say so of cultures, God, or any other individual or group. On this view goodness is "out there" in the realm of things to be discovered. It needn't be "way out there," like goodness in some cosmic sense or goodness for the universe at large. We're

37. Meta-Ethics

The ethical claims listed above are all general in the sense that they make claims that are intended to hold for lots of people in lots of situations. But not all of these claims are "absolute," where this means something like "no exceptions allowed." Taking honesty to be a virtue doesn't necessarily mean that it would be wrong to mislead the Nazi SS officer about the Jews hiding in your attic. And ethical claims needn't be so general. For instance, "It would be wrong to torture Laura's puppy after we go the movies on Friday," is a pretty specific claim about particular things. But it is still a candidate for being an ethical truth. So there might be plenty of ethical truths even if there aren't any true absolute universal ethical generalizations.

Hopefully we are now clear about what sorts of claims are candidates for ethical truth. Now, what would it mean for any claim like those listed above to be an ethical truth? Ordinarily, when a claim is true there is some fact out there in the world somewhere that makes it true. If it's true that Russ' favorite bike has 20 speeds, then what makes this claim true is that there is a certain object in the world that is Russ' favorite bike and it has 20 speeds. So, one pretty straightforward proposal is that if there are ethical truths, then there are corresponding facts in the world that make them true. These facts needn't involve concrete physical objects like my favorite bike. We often attribute rightness or wrongness to kinds of actions, for instance. So it might be that certain kinds of action, like torturing innocent puppies just for fun, have ethical properties like wrongness. Likewise, certain social institutions could have ethical properties of justice or injustice, characteristics of personalities could be virtuous or vicious.

We are narrowing in on a way to understand a view we will

call *ethical realism.* Ethical realism is the view that there are ethical truths and that they are made true by facts independent of anyone's say so, will, or sentiment. These facts will be the *truth-makers* for ethical truths. We will examine a few realist ethical theories of right action in the next chapter. For any realist ethical theory, we will want some account of what makes the theory true, if it is true. This can be given in terms of a theory of objective value. Utilitarianism, for instance, says that right action is action that maximizes overall happiness. This realist ethical theory is based on a view about objective value. Namely that happiness has value (objectively, independent of how much we might like it).

We live in an ethically skeptical age. Many people fail to recognize ethical realism as a serious contender when they think about whether there are ethical truths and what could make them true. Usually, when people think there are ethical truths at all, they take them to be made true by people or God, rather than objective value. We will call this view *ethical conventionalism.* This view makes ethical truth a matter of convention. We can point to familiar examples of things that are true and made true by convention. It's against the law to drive drunk, and what makes this true is an act of the legislature. This is a pretty formal convention. But there are also less formal conventional truths. It's rude to spit in public, but what makes this true is a much less formal, generally unspoken social convention. So, one view about ethical truths is that they are like truths of etiquette or law. Perhaps morality is something like a really serious variety of politeness. Moral truths, on this view, are more or less formal social conventions, made true by the will, say so, or sentiment of a social group and holding only relative to that social group.

What we are describing is a view commonly called *Moral Relativism.* This is one of the more popular versions of ethical conventionalism.

Conventionalist ethical views needn't make morality relative to social groups or the say so of people though. Another very

popular conventionalist view of morality is *Divine Command Theory* (DCT) which holds that there are moral truths and they are made true by the will or command of God. Morality is not relative to social groups according to DCT. It is absolute and holds everywhere for all people. But DCT is still a variety of conventionalism because it makes what is good or bad a matter of convention, just God's rather than ours.

Conventionalist views of ethics, either DCT or Moral Relativism, are far and away the most popular among the philosophically untutored. Conventionalism is also the most roundly rejected view about the nature of ethics among philosophers. Much of this chapter will be devoted to making it clear why Moral Relativism and DCT are both, well, horrible views about ethical truth. Religious believers and non-believers alike have better options.

There is one further meta-ethical position to introduce before we consider our options in greater detail. An alternative to realism and conventionalism is that there are no ethical truths at all.

We'll call this view *ethical subjectivism*. You might recall David Hume holding a view like this. Many others, including the Logical Positivists, have endorsed something like ethical subjectivism. The sentences on our list above certainly look like the sorts of sentences that could be true or false. The sentence "Honesty is a virtue" seems to be a simple subject predicate sentence that asserts something about honesty. But according to the subjectivist, this isn't the sort of sentence that could be true or false because there is no such property as being a virtue. In fact, another way to understand ethical subjectivism is as the view that there are no ethical properties. If there are no ethical properties, then being virtuous can't be a property of honesty. Likewise, we can't attribute goodness to paying your taxes or wrongness to torturing puppies according to ethical subjectivism because there is no property of goodness or wrongness to attribute.

We might be tempted to say that if there are no ethical truths

then it would be ethically OK to do whatever we want. But, perhaps surprisingly, ethical subjectivism denies this too since there is no property of being ethically OK to attribute to whatever we want to do. Subjectivism doesn't settle any questions about what we should or shouldn't do. It is just the view that there are no ethical properties and hence there are no ethical truths.

Subjectivists like Hume don't deny that we have ethical sentiments. We feel indignant at the thought of torturing puppies, for instance. A subjectivist can readily grant this and take our moral and ethical talk to be ways of displaying our moral sentiments. This view is sometimes called "yea-booism" since it takes sentences that look like ethical claims to in fact be displays of ethical sentiment. So, the real meaning of "It's wrong to torture innocent puppies" is something more like "Boo, puppy torture!" Exclamations like this can display our feelings. But exclamations like "Boo, puppy torture!" or "Yea, go team go!" just aren't the sorts of sentences that can be true or false. They don't assert anything. We can feel just terrible about puppy torture without puppy torture itself having any kind of ethical property.

In this section we have introduced three general meta-ethical positions:

- Realism is the view that there are ethical truths and they are made true by something other than convention.
- Conventionalism is the view that there are ethical truths and their truth is a matter of convention (God's in the case of DCT, people's conventions in the case of Moral Relativism).
- Subjectivism is the view that there are no ethical truths, only subjective ethical sentiments.

It should be clear that these three meta-ethical positions cover all the logical possibilities. In the remainder of this chapter we will take up some evaluation of these positions. As we've already

mentioned, conventionalist ethical theories are pretty uniformly rejected by philosophers and we'll want to get clear on why. Subjectivism is a contender, but a puzzling one. We will say a bit more about challenges for subjectivism. In rejecting conventionalism and raising problems for subjectivism, we build a case for ethical realism. Of course this meta-ethical position may face its own concerns and we needn't settle the score between subjectivism and realism here. But for reasons that will become clear soon enough, we will need to get conventionalism out of the way before we take up our inquiry into normative ethical theories in the next chapter.

Against Conventionalism

There are many variations on conventional ethical theory depending on who gets to say what's right or wrong for whom. What they all have in common is that these theories make right and wrong a matter of somebody's authority or some group's authority. Since it is generally actions that are commanded, we will conduct this discussion in terms of right and wrong action. Things would go pretty much the same if we conducted our inquiry in terms of virtue and vice or good and bad more generally. To keep things simple, we will just discuss the two views we've already mentioned: DCT (Divine Command Theory) which makes right and wrong a matter of God's say so, and societal Moral Relativism that makes right or wrong relative to a society's say so.

According to DCT, what is right is right simply because God commands it. This view makes ethics easy, so long as we can be sure we know what God commands. If we can somehow be confident about that, ethics requires no critical thinking, just total obedience. We had a much earlier encounter with DCT in our discussion of Plato's dialogue, *Euthyphro*. In that

dialogue Socrates points towards the classic and still cogent objection to DCT. The central problem for DCT is that it makes ethics completely arbitrary. In principle, God could command that anything be right. God could command that we torture puppies, commit genocide, and treat children like livestock. According to DCT, if God does command these things, then they are right, end of story. In fact, many people have sincerely taken God to have commanded these things (perhaps except for puppy torture). However, hopefully, the idea that any of these things could be morally right strikes you as absurd. In spite of our occasionally differing ethical opinions, ethics does seem to be systematic and coherent. Right and wrong are not completely arbitrary. It seems at least that there is some reasoned systematicity to our ethical opinions in spite of the differences we sometimes arrive at. If this is right, then we should reject any meta-ethical view that makes ethics completely arbitrary. And this means rejecting the view that right and wrong is simply a matter of God's command.

The religious believer has better meta-ethical options than DCT. When I present students with the knock down objection to DCT just given, it's not uncommon for someone to object that God would never command us to torture innocent puppies because God is good. I think this is exactly the right response for a believer to offer. But this response is not a defense of DCT. Any believer that makes this move is joining Socrates in rejecting DCT and taking God to command what is good because it is good. If God is essentially good, then what is right is not made right merely by his command. Rather he commands what he commands because of his goodness. When the religious believer takes God's goodness to be what is ethically fundamental he abandons conventionalist meta-ethics in favor of a kind of theological ethical realism. Of course, the challenge of understanding God's good nature remains.

People whose ethical opinions are not guided by religious faith have a very unfortunate tendency to retain the

conventionalist authority-based view of the nature of ethical truth. The result, most frequently, is some variety of Moral Relativism. Perhaps the shift to Moral Relativism is based on the assumption that if there is no God to decide what's right and wrong, then it must be people who get to decide right from wrong. The idea that ethics might be a matter of inquiry and discovery rather than authority and command seldom gets a foothold without some structured philosophical critical thinking. Descartes' vision of shaking off the shackles of authority and thinking freely is far from fruition in our culturally dominant way of thinking about morality.

Let's take societal Moral Relativism to be the view that what is right relative to a society is whatever is deemed right by that society. We could ask for a few clarifications. In particular, it would be good to know what counts as a society and what it takes for a society to deem something right. In the broadest sense, we might take any social group to constitute a society, though I don't think anyone is a chess club moral relativist or a garden society moral relativist. People are much more inclined to take culture to identify the social groups relevant to morality. And this sounds appealing given that moral traditions are often incorporated into cultural traditions. Keep in mind, though, that ethics is about what moral opinions are best, not what moral opinions are in fact held by people or how they come to be held by people. While most of us are pretty likely to inherit our moral opinions from the dominant traditions in our culture, being entrenched by culture might not be the best guide to what is good. Given this, we might ask why it is culture that gets to decide right and wrong rather than the chess club or the garden society. Moral Relativism seems to suffer a kind of arbitrariness even at the level of selecting the groups to which right and wrong are supposed to be relative.

Next, what is it for a group to deem something right or wrong? As we are culturally engrained to think egalitarianism is a good thing, most of us would probably say that a group deems

something right when a solid majority of its members deem it right. But why not take a group to deem something to be right with the strongest and most aggressive member of the group deems it right? This is how things work with gangs and outlaw militias. If right and wrong are merely matters of convention, why should we favor egalitarian democratic say so over gangland style strongman say so? Note that it won't do to appeal to values independent of the say so of groups here, since Moral Relativism denies the existence of any value independent of group say so. It appears that a further element of worrisome arbitrariness lurks just in the attempt to formulate a plausible version of Moral Relativism.

Whatever version of Moral Relativism we lump for, the problems will be basically the same. Because Moral Relativism grounds right and wrong in authority, it suffers the same central problem as DCT. The commands of people can be just as arbitrary as the commands of any god. Anything can be right relative to a culture. All it takes is for the culture to deem it right. So if a culture deems it right to cut the genitals of young girls without regard to their consent, then, according to cultural Moral Relativism, this is right relative to that culture. Should this example seem at all ethnocentric, let's add another. If a culture deems it good for women to walk around all day in shoes that wreck their feet, then, according to cultural Moral Relativism, this is good relative to that culture. If neither of these examples strikes you as morally absurd, then consider racism, genocide, terrorism, or exhausting natural resources leaving future generations to suffer and die off. According to Moral Relativism, all it takes for any of these things to be right relative to a culture is for that culture to deem it right. As ethical theories go, Moral Relativism begins to look like a bit of a train wreck. Yet all we have done here is reason very straightforwardly and deductively from what Moral Relativism says.

The arbitrariness of Moral Relativism leads directly to the central and most compelling objection to the view. But there is

more to consider including dispelling some myths that seem to speak in its favor. Many would endorse some version of Moral Relativism on the grounds that it seems to support tolerance and respect for societies with differing moral views. Moral Relativism seems to be a view that allows for different societies to embrace different moral standards that are right relative to the respective societies. Moral Relativism rejects the notion that the moral standards of one society could be objectively correct. This line of thought has led many who value cultural diversity and tolerance to embrace Moral Relativism. But this is a mistake. Moral Relativism does not entail that we should be tolerant of diversity. Moral Relativism entails that we should be tolerant of diversity if and only if our group deems tolerance of diversity to be a good thing. If a group deems intolerance to be good, then, according to Moral Relativism, intolerance is good relative to that group. Since goodness is relativized to groups, our view that tolerance and respect for diversity is good fails to provide the intolerant group with any grounds for reconsidering its intolerance. Moral Relativism thus turns out to be a deeply conservative view in the sense that it undermines all possible reasons for changing our moral outlook. Moral Relativism is a view that gives the dominant racist culture moral standing and further denies us any reasonable grounds for arguing against the intolerance of the dominant racist culture. We who value tolerance and respect for diverse individuals or groups would do much better to endorse tolerance and respect as objective realist ethical values than to endorse Moral Relativism.

A further strong argument against Moral Relativism is the argument from change. Sometimes our view about the moral status of some practice changes. A person might, for instance, think that eating meat is morally unproblematic at one time and then become convinced that animals deserve some kind of moral regard that speaks against eating them. When moral views change in this fashion, people do not merely drop one moral belief in favor of another. Typically, they also hold that

their previous moral views were mistaken. They take themselves to have discovered something new about what is morally right. Likewise, when the prevalent moral belief in a society undergoes a significant change, as in the civil rights movement, we are inclined to see this as a change for the better. Moral Relativism has no problem with changes in moral standards. But the relativist cannot account for any changes in our moral beliefs as being changes for the better. This is because the Moral Relativism recognizes no independent standard of goodness against which the new moral opinions can be judged to be better than the old moral opinions.

A closely related problem for Moral Relativism is the moral reformer's dilemma. We recognize a few remarkable individuals as moral reformers, people who, we think, improved the moral condition of their society in some way. Common examples might include the Buddha, Jesus, Ghandi, or Martin Luther King Jr. While the relativist can allow that these individuals changed the moral views of their societies, none can be said to have changed their societies for the better. Again, this is because the societal moral relativist recognizes no standard of moral goodness independent of what is accepted in a society according to which a society that changes can be said to change for the better. The relativist is committed to taking the most overt forms of racism to be right relative to pre-civil rights American society and wrong relative to post-civil rights American society. But since standards of goodness are determined by the prevalent views in a society, there is no standard of goodness to appeal to in judging that the change our society underwent in the civil rights movement was a change for the better. According to societal Moral Relativism, anyone who takes Martin Luther King to have improved American society by leading it to reject many forms of racism is just mistaken about the nature of ethical truth.

Relativism and the Social Sciences

The social sciences are in the business of trying to better understand and explain the diversity of cultural practices and world views. But in describing culturally based beliefs about what is right or wrong, they are not defending ethical claims about what is right or wrong. The social sciences are often concerned with what people in different cultures believe is right or wrong. And social scientists will often discuss a kind of descriptive cultural relativism in explaining how what is deemed good or bad in various cultures is relative to their respective values and traditions. But the question of what *is* good or bad remains a question for *ethics*.

Suspending judgment is methodologically important for understanding. This is just as true in philosophy as it is in sociology or anthropology. We suspend judgment at the stage of trying to understand a new view. Only once we have a clear understanding can we then turn to critical evaluation. The social sciences are out to understand cultural practices and perspectives and suspending judgment is essential to doing this well. So guarding against ethnocentrism is important when an anthropologist investigates cultures that are different from her own. But the methodological importance of suspending judgment for the sake of better understanding is not a permanent obstacle to critical evaluation of the moral points of view transmitted through culture. Ethics, unlike sociology or anthropology, is a fundamentally normative discipline. Its goal is to evaluate moral views and try to see which is most reasonable in light of the kinds of ethical evidence and argument we can uncover. Here we benefit from the social sciences and the understanding they produce of the moral perspectives of different cultures.

When we take up ethics and critically evaluate moral opinions, we are moving beyond the suspension of judgment.

In ethics our goal is to better understand which moral opinions are reasonable and which aren't. But our ethical judgments are to be grounded on ethical reasons. It remains just as important that we avoid ethnocentrism in evaluating moral views. Criticizing a practice that is morally accepted in another culture because it is not in line with in our own culturally based values is simply a non-starter as an argument in ethics. If we have some reason for thinking that an evaluation of a moral opinion is based on some culturally loaded value or bias, then to that degree we have a good reason to discredit that evaluation.

Lots of people find societal Moral Relativism appealing as a means of conflict avoidance. It is a way for everyone to feel that they have things right. But, to engage in a bit of social science, relativism about morality seems plausible only in comfortably decadent cultures. Nobody buys Moral Relativism once someone starts shooting. When you don't have the option of avoiding conflict, that there is a difference between just and unjust, right and wrong, is often too starkly apparent to ignore. Given this, we should worry that Moral Relativism as a means of conflict avoidance is really a lazy and cowardly way for the comfortable and complacent to avoid addressing important issues.

There should not be so much to fear in investigating ethical issues. When we sit down to formulate and evaluate ethical arguments, it's not really about who is right or who gets to have their way. Like any other kind of inquiry, it's really about looking into issues and trying to reason well. Rational inquiry done well doesn't have to include unpleasant conflict, but it does hold out some hope for resolving conflicts reasonably. In ethics we put an argument for a view about what is right or wrong on the table and talk about the quality of the argument. Where the argument came from is not what is at issue at this point. Neither is who likes or dislikes the conclusion. All that is at issue is whether or not the premises of the argument should

be accepted, and whether or not the conclusion follows from the premises. Making carefully reasoned judgments about ethical views is not the same thing as condemning or seeking recourse against those who hold them. Careful inquiry into what is good, right, or just is an essential precursor to effectively fighting for social justice. But in the context of inquiry, we are not joining the battle and to conflate these two activities is very likely to result in doing both of them badly.

Philosophy is only concerned with whether or not good reasons can be given for accepting or rejecting positions and opinions. Free and open inquiry, inquiry that employs as many diverse perspectives as possible, provides the only method we have for identifying and filtering out culturally based biases. Bringing a righteous battle to inquiry can only silence voices whose inclusion would be valuable.

For yet another compelling line of argument against Moral Relativism, see Paul Boghossian's piece, "The Maze of Moral Relativism." Boghossian argues that attempts to relativize morality undermine the normativity of moral beliefs altogether and so ultimately collapse into nihilism, the view that nothing matters, nothing is good. If you prefer to listen, here's a Philosophy Bites podcast in which Boghossian explains his line of argument.

If ethics is a matter of authority as both DCT and Moral Relativism would have it, then there is no inquiry to engage in beyond figuring out what the relevant authority says. This would make ethics a singularly boring topic to look into. But we will find quite a few interesting things to say about plausible normative ethical theories. So we might take our inquiry into normative ethics in the next chapter to constitute one further argument against conventionalist approaches to ethics. Ethics just isn't as dull as conventionalism would have it. Before we get there, we need to address subjectivism, the view that there are no ethical truths, or no ethical properties.

Against subjectivism

Here, I want to discuss just one consideration that I think speaks for a realist view of ethics over the subjectivist view. We seem to reason about ethics quite a lot. We don't just express ethical sentiments, but we incorporate ethical expressions into complicated strings of expressions that look an awful lot like arguments. People who think the death penalty is wrong don't just say "The death penalty is wrong." Sometimes at least, they also say things like, "The death penalty is wrong because it involves the killing of a person and it's wrong to kill a person." It certainly seems like what is offered here is an argument. And we commonly evaluate such expressions as if they were arguments. But if the subjectivist is right, then whatever the opponent of the death penalty offers with this expression, it isn't an argument. That's because, as we learned in Chapter 2, an argument consists of a series of claims that admit of truth or falsity. In order to be a part of an argument (in order to be a premise or a conclusion) a sentence has to be a statement that makes some claim about how things are (and therefore is capable of being true or false). But the subjectivist who follows Hume in taking moral sentences like "murder is wrong" to be mere expressions of sentiment, equivalent in this case to "Boo, murder!" denies that such sentences make claims that admit of truth or falsity. Subjectivism would thus have it that the apparent line of reasoning against the death penalty mentioned above should really be taken to express something like this:

1. The death penalty is the killing of a person.
2. Boo, killing persons.
3. So, boo the death penalty.

38. Feminist Ethics

The idea that there might be a single universal and absolute criterion of morally right action strikes many who value cultural diversity as highly problematic. But lest we abandon monist approaches to ethical theory too quickly, we should note that the standards of right action offered by both the utilitarian and the Kantian are highly abstract and for this reason they are quite compatible with a rich range of diversity in more specific derivative guidelines for action. In fact, lots of cultural diversity can be explained in terms of more broadly shared underlying moral values. Eating the dead may be seen as a way of honoring them in one culture, but be considered a sacrilege in another culture. Both of these diverse practices can be seen as diverse ways of expressing respect for persons. The difference between cultures in this case is not really a difference of fundamental moral values, but a difference in how these are to be expressed. Similarly we consider infanticide morally wrong while other cultures facing more difficult environmental pressures may practice it routinely. What may seem like conflicting moral standards at this more specific derivative level might instead be understood as differing ways of maximizing happiness that are appropriate for the starkly different circumstances that the respective cultures must deal with. So absolutist, universalizing, monist ethical theories turn out to be considerably more accommodating of cultural diversity than we might have thought at first. Still, they may not be flexible enough.

It might be that some cultures value respect for persons over happiness while others value happiness at the expense of respect for persons and others yet value community or kinship relations more than happiness or respect for individual persons. That is, we might find conflicts in the most basic or fundamental moral values upheld by diverse cultures. How

can ethical theory account for this without begging questions against one set of cultural values or another?

Recall that the ethical monist is out to discover a single rationally defensible moral truth that is grounded in a single kind of moral value. In discussing monist ethical theories I insisted that you can't be both a utilitarian and a Kantian respect-for-persons theorist. This is because these theories offer logically incompatible principles of morally right action. There will be actions (like harvesting the healthy patient's organs in the simple versions) that one theory will deem to be right and the other will deem to be wrong. So, you can't coherently hold both a utilitarian principle of right action and a Kantain principle of right action to be true. If the principles disagree on even a few cases, they can't both be true. But let's set principles aside for a moment. I'm not suggesting we be unprincipled, I just want us to focus on the underlying moral values without worrying about truths that might be based on them. There is nothing logically incoherent about taking happiness and respect for persons to both be good in fundamental ways. And there may be other plausible candidates for fundamental goodness. Happiness and respect were just the ones that got most of the attention in the 18th and 19th century. Since then, feminist philosophers have argued that we should recognize a fundamental kind of value in caring relationships.

Environmental ethicists have argued that we should recognize a fundamental kind of value in the natural world. Hindus and Buddhists have long suggested that there is a kind of fundamental value in consciousness.

Perhaps this short list is long enough. Or perhaps it is already too long. A moral value is only fundamental if it can't be explained and supported in terms of some other fundamental value. So if caring relationships matter just because they bring happiness to human lives, then we already have this kind of value covered when we recognize happiness as a kind of fundamental value.

But it is not at all clear that happiness fully explains the value of caring relationships. There are issues to explore here and feminist philosophers are just starting to map out this terrain. In any case, kinds of fundamental value might be rare, but still plural.

So what should ethical theory say about cultures that differ in the fundamental values that shape their customs and codes? Monist approaches to ethical theory would insist that we pick winners in this kind of situation. But should we? Certainly, in some cases we should. The fundamental values of Nazi culture were racist through and through. Good ethical theory should not be accommodating this kind of cultural diversity at all. Recall that our most compelling argument against Moral Relativism was that it is committed to accepting that racism is right relative to racist societies and our condemnation of racism has no more moral force than their endorsement of it.

39. Attribution

The material in this chapter is adapted from previously published content:

Payne, R.W. (2015). *An Introduction to Philosophy*. BCcampus. Sections 9.97-104, 10.108-114, 10.116 https://open.umn.edu/opentextbooks/textbooks/598 CC BY NC.

Versioning History

This page provides a record of changes made to this textbook. Each set of edits is acknowledged with a 0.01 increase in the version number. The exported files for this ebook reflect the most recent version.

If you find an error in this ebook, please send an email to library@sait.ca.

Version	Date	Details
1.0	September 202`1	Textbook is published